

u/Mods and u/Rockers:

Constructing community identity through moderation practices in location-based subreddits

by

Grace Wilson Chiarello Barnett

(Under the Direction of Karin Assmann)

ABSTRACT

This study examines the relationship between moderation practices and community identity in location-based subreddits, exploring how these dynamics influence discourse and governance in digital spaces. Through a qualitative approach analyzing thousands of posts across multiple location-based subreddits, this research finds that moderation practices do not significantly alter the topics discussed but instead provide a channel through which users negotiate relevance and community norms. Community identity, in turn, plays a crucial role in informing moderation practices, as users collectively challenge, reinterpret, and reinforce guidelines in ways that reflect local culture and digital engagement norms. The study highlights the interplay between platform policies, user behaviors, and digital governance—revealing that subreddit communities function through an ongoing negotiation between moderation and user agency. These findings contribute to broader discussions on online community formation, digital moderation, and the evolving role of user participation in shaping digital public spheres.

INDEX WORDS: Content moderation, community identity, digital culture, trust and safety, Reddit

U/MODS AND U/ROCKERS:
CONSTRUCTING COMMUNITY IDENTITY THROUGH MODERATION PRACTICES IN
LOCATION-BASED SUBREDDITS

by

Grace Wilson Chiarello Barnett
B.A., The University of Georgia, 2021

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment
of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2025

©

Grace Wilson Chiarello Barnett

All Rights Reserved

U/MODS AND U/ROCKERS
CONSTRUCTING COMMUNITY IDENTITY THROUGH MODERATION PRACTICES IN
LOCATION-BASED SUBREDDITS

by

Grace Wilson Chiarello Barnett

Major Professor: Karin Assmann

Committee: Jonathan Peters
James Hamilton

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2025

DEDICATION

For my grandparents.

ACKNOWLEDGEMENTS

Thank you to Dr. Assmann, who's midnight comments are the reason this paper exists.

To my dog, Joan, for waiting patiently at the door to be taken out while I stood at my desk

finishing just one last thought.

To my best friends in the whole world: Katya Abuhoff, Waled Atwa, Lacey Coleman and Anna

Wilson. Thank you for listening.

To my mom and my sister, for always being there.

And to Athens, Georgia.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	V
LIST OF TABLES	VII
LIST OF FIGURES	VII
CHAPTER	
1 Introduction	1
2 Literature Review	6
Community	6
Online governance	18
Community, content moderation, and Reddit	29
3 Methodology	37
4 Findings	52
Community guidelines	56
Moderating moderation	76
Content of discourse	87
5 Discussion	102
Moderating and shaping online discourse	103
Community identity and moderation practices	109
6 Conclusion	123
REFERENCES	136

LIST OF TABLES

	Page
Table 1: Data selection metrics	41
Table 2: Total size and date range of dataset	45
Table 3: Classification of codes	48
Table 4: Content of subreddit guidelines	57

LIST OF FIGURES

	Page
Figure 1: Subreddit layout	54
Figure 2: Guideline classification	57
Figure 3: Political meme shared in r/Philadelphia	95

CHAPTER 1

INTRODUCTION

On January 7th, 2025, Mark Zuckerberg, CEO of Meta, announced the company's decision to eliminate fact-checking operations on Facebook (Chan et al., 2025). Framed as a move toward more neutral moderation practices, Zuckerberg revealed a new strategy: a shift from the use of professional third-party fact checkers to reliance upon a "Community Notes" model. This model, based on user-driven content evaluation, is similar to the approach used by platforms like X (formerly Twitter) where fact-checkers were removed following Elon Musk's acquisition of the platform. According to X's site, "Community Notes aims to create a better-informed world, by empowering people on X to collaboratively add helpful notes to posts that might be misleading" (X, 2024). This approach provides users with control over the content that appears on the platform through features like liking, commenting, and flagging posts.

Meta's Zuckerberg cited rising incidents of political bias from professional fact-checkers as justification for the shift, admitting in a follow up interview that misinformation would likely increase but was a necessary tradeoff to maintain a space for inclusive discussion. Meta's site claims that, after a short developmental period, Meta employees will not write any community notes—only the users (Meta, 2025). While further specifics about the new initiative are currently limited, it is clear that the structure will closely mirror that of X's community notes—most notably in its reliance on what X calls a "diversity of perspectives". Users are assessed according to their engagement history on the platform to determine their perspective, and multiple users of varying perspectives must all agree on the post in question. Meta's remaining moderation at the

platform level seek to address sensitive and harmful content but not misinformation (Community Notes, n.d.).

The interplay between external forces and the nature of content moderation in digital spaces raises significant questions about the role of digital platforms in regulating speech. Reddit, dubbed “the nicest swamp on the Internet, has emerged as a significant platform for participatory media consumption and production (LaFrance, 2025). Unlike other social media platforms, Reddit’s structure allows for user-driven organic discussions and content sharing, making it an ideal space to observe how digital platforms can fill gaps in local and community-based information. In 2024, Reddit was the eighth most-known social media platform in the United States (Kunst, 2024). Its user base has doubled in just five years, coming in at 11% of U.S. social media users in 2019 and 22% in 2024 (Ceci, 2024). Further, almost half of its users are aged 18-29 years old (Sidoti & Dawson, 2024). Reddit stands apart from many other popular social media sites in the way its members can engage in deliberate, self-selected communities of interest to form organically, rather than through algorithmic curation. This decentralization lends itself to focused, community-driven discourse.

These self-selected, topic-oriented groups called subreddits range widely in subject matter. Although Reddit’s user base is similar to that of platforms like Facebook and X with 50.7 million members (Dixon, 2024), academic engagement with Reddit has been relatively limited. Much research about Reddit focuses on the more taboo corners of the site— drug usage, politics, and mental health being among the most popular (Proferes et al., 2021). Many of its characteristics make it a compelling data source for researchers. All that’s needed to make an account is an email and a username, which is auto-generated by Reddit and can be changed by the user if desired. Those who choose not to make an account can still browse completely

unhindered. Most users who do make an account do not use their own name or otherwise identifying information. The creation of burner accounts is common as one can make many different accounts with all the same information, allowing the creation of accounts for very specific purposes (Hartzell et al., 2021). This allows people to speak freely in a manner they might otherwise not and researchers to act as flies on the wall— providing completely unfettered answers to whatever question they are asking. Further, Reddit’s open-access dataset is an asset to researchers. However, focusing solely on the content of the posts generated on Reddit ignores the equally valuable insights that can be gained from the processes involved in creating and interacting with them. The dynamics of content distribution on Reddit reveal fundamental aspects of how communication practices are evolving on digital platforms, making it essential to study both the content and the mechanisms behind its dissemination.

While Reddit communities are sometimes extreme or controversial, many subreddits form around mundane topics. Among these are location-based subreddits, which are organized around specific geographic locations ranging from entire states to specific neighborhoods. These subreddits function as hyper-localized digital spaces where users exchange a wide array of information, from daily events and public services to more nuanced discussions about community issues. Sprinkled in between are inside jokes, references to local culture, and discussions of shared experiences. While new media scholars have certainly done work that examines how these interactions on social media sites foster community (Lingel, 2021; Preece, 2000), location-based subreddits have not been studied.

Reddit possesses unique features that set it apart from competitors like Facebook and X. Reddit’s technical infrastructure, particularly its system of upvotes, downvotes, and comment threads, allows for rapid dissemination and simultaneous community vetting of information.

Further, its open-source code means that its construction is a result of iterative processes enacted by both community members and administrators (C. Miller, 2015). While Reddit does host ads, its primary source of revenue comes from its Gold membership. Gold members are granted a completely ad-free experience. This is a significant deviation from many other traditional and new media sources, and its community centricity allows it to cater better to the interests of its userbase. Further, this is a model that is mutually beneficial to all who use Reddit. Those who can and will pay for a Gold membership support the continued existence of the site for those who don't pay. Those who don't pay are still granted the same access to the platform as those who do.

At its core, Reddit is a community-driven platform that encourages users to adapt and shape their own experiences within a system guided by collective control. Unlike more centralized media forms, Reddit allows for the creation and enforcement of rules by its user communities, which provides a unique lens through which to examine how content moderation and community management shape the flow of information and discourse. The interaction between these user-driven governance structures and the information needs of community members is a key focus of this study, as it reveals how communities self-organize and maintain order while meeting diverse informational demands.

Given that these subreddits often operate alongside established local and national media, they represent a unique blend of grassroots and formal information-sharing practices. By providing decentralized spaces for users to engage with hyper-localized content, location-based subreddits offer an alternative to more traditional forms of media. As such, this study investigates how content moderation—both user-driven and platform-enforced—shapes interactions within these subreddits, influencing not only the flow of information but also the evolution of community norms. By situating these questions within the context of participatory

digital environments, this inquiry aims to illuminate the mechanisms that allow Reddit's decentralized, user-centric model to both facilitate community-building and manage information flow in ways that differ from more conventional media systems.

Through a qualitative content analysis this study analyzes four location-based subreddits—selected to reflect varying levels of moderation and user size, that represent a wide range of community-specific factors and diverse cultural expectations and social norms. Key areas include the types of information shared, the dynamics of user interaction, and the role (and reception of) moderators in shaping community norms. To fully grasp the impact of content moderation, and the forces that impact its deployment, it is essential to explore how these digital communities are structured and how they make sense of themselves through their participatory processes. This is especially true as these platforms become the primary source of information procurement and social interaction for many people.

As digital platforms continue to play an increasingly central role in public life, understanding how they mediate community information-sharing practices through governance is critical. Reddit, with its decentralized structure and user-driven methods, offers a valuable case study for examining these dynamics. By analyzing the interplay between moderation, community-building, and information sharing, this study aims to shed light on the broader reasons for and implications of the outsourcing of governance labor to users.

In the following sections, I provide a review of the existing literature on content moderation, freedom of speech on social media platforms, and information-seeking regimes on digital platforms. This will set the foundation for exploring how these concepts intersect within the context of Reddit's unique model, where self-regulation, user participation, and collective control shape the platform's governance and impact on the information ecosystem.

CHAPTER 2

LITERATURE REVIEW

This literature review examines key theoretical frameworks and empirical studies relevant to understanding the dynamics of online communities on social media platforms, particularly within the context of Reddit. It explores understanding of how communities have been formed and maintained throughout history, expanding these concepts and applying them within the context of online communities in the digital age. Communities are shaped not only by the behaviors of their users and interactions among them but also by broader systems of platform culture and governance practices that influence their structure and evolution. To understand the complexities of digital spaces like Reddit, I examine existing research on community formation, content moderation, identity and community construction on Reddit. Each of these elements plays a pivotal role in shaping these online environments.

Community

Community, most familiarly, is formed through shared physical spaces (Small & Adler, 2019). However, communities can also form around social and cultural measures: shared interests, belief systems, education level, and economic standing to name a few (Buckingham et al., 2018; Lauer et al., 2024). Community can also be achieved through legislative means, by the formation of policy that either defines its boundaries or supports its development (Vos, 2005). Community formation has also been perpetuated through the affordance of technological advancements (Alsaleh, 2024; Fibrianto & Yuniar, 2019). Conversely, loss of community has

long been proposed as a result of societal ills, its development and maintenance a solution to those same ills (Willson, 2006).

Sociologist Ferdinand Tönnies (1887) distinguishes between two forms of social organization: *Gemeinschaft* (community) and *Gesellschaft* (society). In *Gemeinschaft*, social bonds are based on close personal relationships, shared values, traditions, and a sense of belonging. These communities are formed organically through kinship, locality, or deep interpersonal ties. They emphasize mutual dependence and collective well-being. *Gesellschaft* represents a structured, impersonal form of association. Individuals engage based on rational self-interest, formal agreements, and institutional regulations. While *Gemeinschaft* fosters an intrinsic sense of identity and cohesion, *Gesellschaft* is marked by the necessity of rules, policies, and systems to maintain order.

But communities are complex organisms that exist at various levels of organization according to a wide range of characteristics. More conservative social scientists utilize a biological definition of community, arguing that community has no social aspect and is purely geographical in nature (R. E. Park, 1936). Others argue for the reliance on alternative characteristics for the formation of meaningful communities (Gergen, 2003; Meyrowitz, 1985), such as the democratization of technologies that may potentially create new spaces within which individuals engage and form bonds (Antonucci et al., 2017; Horrigan, 2001; Small & Adler, 2019). Because of this, conceptualizing the delineation between “physical” community and “virtual” community is a necessary component of communications studies (Kavanaugh, 2014). Virtual communities redefine the boundaries of traditional communities by extending social interaction beyond space and time, offering new modes of connection otherwise unrestricted by limitations of physical proximity (Coeckelbergh & Reader, 2023; Katz & Rice, 2002;

Kolozsvari, 2012). However, virtual and physical worlds do not exist completely separately. Instead, it is better to conceptualize both as ideal types that possess certain characteristics but are ultimately part of the same experience (Katz et al., 2004). Any stringent delineation between two disregards the Internet's potential for social relations to be maintained offline by positioning the Internet as an entirely different social system (Hampton, 2004).

Instead, we must consider the formation of hybrid communities that blend virtual and physical elements (Franco & Birenboim, 2024). Research indicates that interaction in online communities facilitates offline action that extend into all realms of life— personally and professionally, at both the individual and societal level (Babar, 2020; Oksanen et al., 2024). Moreover, studies have shown that offline interactions among online community members can strengthen online participation— suggesting a mutualistic enhancement between virtual and physical engagements (Tamí-Maury et al., 2017).

Defining community in the digital age

Technological advancements throughout history have facilitated the advancement of communities defined through shared interests, practices and languages. In this way, community transcends geographic boundaries. Anderson (1991) posits the idea of the imagined community. Though they rarely meet in person, members still share a sense of belonging through common symbols, narratives, and practices. From nations to religious groups to political parties, these groups gain strength through the development of faster transportation, the expansion of literacy facilitated by the printing press, and other advancements. Improved communication frameworks enhance people's ability to engage in shared practices and connect with others who share their identities. It is through this shared language that identity is formed. Online communities are imagined in the sense that their members, despite not physically interacting, develop a shared

sense of belonging through digital symbols, narratives, and linguistic conventions (Kavoura, 2014). This formation of community occurs through the collective interpretation of meaning, where users engage with content, inside jokes, memes, and platform-specific jargon to establish a cultural framework that defines the group. Much like Anderson's concept of imagined communities in the context of nation-building, these digital spaces rely on the perception of shared identity rather than physical proximity or direct personal relationships. Mediated communities, called fandoms, that arise around a shared interest in media products like bands, movie franchises, and video games are good examples of this (Gray et al., 2017). Putnam (2001) suggests that the shift from interactive, community-based activities to more solitary, screen-based consumption reduced opportunities for social connection and collective participation. The broader cultural and behavioral changes personal technologies facilitated—such as passive consumption of content and the replacement of traditional social gatherings with individualized entertainment—eroded the strength of civic and communal life. But at the time of his original writing, mass media was still largely one-way.

Early understandings of media, such as the transmission model of communication, positioned communication as a one-way process of sending and receiving information—information is disseminated to a passive audience (Shannon & Weaver, 1949). But theorists like Carey (1975) argued for a broader, more culturally embedded perspective. Carey introduced the ritual model of communication, which emphasizes communication as a shared social process rather than a simple exchange of messages. The ritual model conceptualizes communication as an ongoing practice that reinforces collective identity, social norms, and cultural cohesion. Through repeated interactions, shared language, and cultural symbols, communities continuously construct and reinforce their sense of belonging. Digital spaces, transcend their lack of

physicality, by serving as ritualistic arenas where community members continually reaffirm their shared values and social bonds through interaction—a process that is central to understanding the function of online communities in the digital age (Carey, 1983, 2005).

This is further supported when oriented against the current infrastructure of digital spaces—characterized by a high level of participation inherent to modern social media as influenced by the Web 2.0 movement (Wasike, 2011). While Americans may be spending more time physically alone, time spent in front of the screen is no longer a solitary activity (Hargittai, 2022). Thus, modern technology, rather than eroding social connection, has reshaped the way communities are formed and maintained—perhaps providing new opportunities for civic engagement. The internet facilitates communal interaction on a large scale, broadening opportunities for engagement while reshaping the spaces in which collective identities and shared meanings are constructed (Wellman et al., 2002). In 2025, two-thirds of the world's population are active on social media (Petrosyan, 2025). Most of the world is conducting some level of social interaction on the internet. Early critics questioned whether communities built on artificial structures could truly replicate the organic, deeply rooted nature of traditional social bonds (Wellman et al., 2002). Dotson (2017) reiterates the importance of place-based community: “Anyone who has enjoyed a summer day on the front porch with friends and neighbors... would likely not equate it with social media. Try getting a Facebook status update to help move a couch or stay for dinner.” (p. 1)

Dotson distinguishes between what they refer to as "thick" and "thin" communal ties in digital spaces. Thick ties, characteristic of traditional in-person communities, are built on sustained, multidimensional relationships that involve shared physical spaces, mutual obligations, and a depth of social knowledge that extends beyond transactional interactions.

These relationships foster trust, reciprocity, and emotional support, reinforcing a strong sense of communal belonging. In contrast, thin ties—often associated with digital interactions—tend to be more superficial, fleeting, and limited in scope, primarily serving informational or entertainment purposes rather than fostering deep social cohesion. While digital platforms allow for increased connectivity and exposure to broader networks, Dotson argues that they struggle to cultivate the same level of intimacy and investment that traditional communities provide. This perspective challenges the assumption that greater connectivity necessarily translates into stronger communities. While social media enables individuals to maintain an extensive web of acquaintances and access diverse perspectives, it may simultaneously erode the need for local, face-to-face interactions that historically anchored community life. Dotson (2015) contends that online interactions often lack the accountability and long-term commitment inherent to thick social ties, raising questions about whether digital spaces can truly serve as adequate replacements for embodied community experiences.

This lost community thesis argues it is dependent on two distinct meanings of “loss”, often posited at the hands of industrialized mass society (Driskell & Lyon, 2002). The first is psychological, which emphasizes the social interaction dimension and the potential alienation that arises from the loss (Bellah, 1995; Putnam, 2001). The second is place-based, emphasizing weakened identification of a social place (Greer, 2017). Studies have demonstrated that diminished place attachment is linked to reduced participation in local activities, lower social trust, and increased social isolation—reinforcing the idea that spatial detachment negatively impacts social well-being (Rollero & De Piccoli, 2010; Zahnow, 2024). Stronger identification with one’s neighborhood fosters social ties, mental well-being, and civic participation. Shared location is a fundamental aspect of community formation that cannot be discounted. However, it

warrants new conceptualization in the face of participatory digital media. Driven by its democratization— has, in many ways, reshaped how many people form and sustain relationships, often transcending the need for but also reinforcing experiences in shared physical spaces.

A growing body of literature on digital communities suggests, the notion of a “lost community” may be less about disappearance and more about transformation (Lingel, 2021; Scheffer, 2024; W. Wang, 2019). The rise of scholarly interest in social networks and digital spaces underscores their significance in contemporary life. Feeling connected to others despite never meeting, a phenomenon known as socio-mental bonds, has become a widespread and meaningful experience (Chayko, 2002). Online peer communities, for example, have been found to provide support and a sense of belonging among members. In an online community, the Internet becomes the place (Feenberg & Barney, 2004). Some would even see this as a benefit—the internet provides an opportunity to engage in social interactions and experience aspects of social life without many of the challenges or constraints of face-to face communication (Song, 2009). Hampton and Wellman (2001) found that those who used the Internet were overall better neighbors than those who weren’t. They were more likely to know the names of their neighbors, visited the homes of their neighbors more often, and knew neighbors from more far-reaching corners of the neighborhood. Similar studies have found that heavy internet users maintain stronger, more constant bonds with people both locally and at a distance (Purcell, 2006).

In the absence of physical space, identity plays a vital role in virtual communities. This understanding is important both in being able to interact with others successfully and simultaneously an ambiguous feat online (Smith & Kollock, 1999). The structure of digital space allows for people of all interests and identities to engage in communities that tailor specifically to individual needs. Studies have examined the effects of online interaction among LGBTQ+

students seeking connection (R. A. Miller, 2017), as well as those on the autism spectrum trying to improve social relations (Sallafranque-St-Louis & Normand, 2017). In doing so, the way these communities define themselves is attracting new users and maintaining a productive space for interaction among them (Lüders et al., 2022).

Languages of dominance in digital communities

Language has a strong correlation to community identity (Soliman et al., 2019). Third wave sociolinguistics argues that speakers are not passive or stable, rather they act as stylistic agents that tailor community dialects based on identity and self-construction (Eckert, 2018). We can understand subreddit communities are communities of practice—formed and bound by mutual engagement, jointly negotiated enterprise, and shared repertoire (Leuckert & Leuckert, 2020). The formation of these digital communities is deeply tied to linguistic practices, language serves both as a structuring force and a means of identity negotiation (Bucholtz & Hall, 2005). Digital communication platforms amplify these dynamics by fostering distinct linguistic repertoires that mark group belonging and influence discourse norms (Tannen & Trester, 2013). The adoption of platform-specific lexicons, inside jokes, and implicit conversational rules contributes to a shared sense of identity while also serving as a mechanism of exclusion for outsiders (Danescu-Niculescu-Mizil et al., 2013). For example, research on networked multilingualism highlights how linguistic variation in digital spaces reflects users' alignment with particular community values and identities (Androutsopoulos, 2013). This phenomenon is further evident in the discursive power of memes and other forms of internet vernacular, which create intertextual layers of meaning that reinforce in-group solidarity (Wiggins, 2019).

When a social identity is engaged, a shared understanding of group attributes fosters uniform behavior and collective purpose among those who identify with that group, often leading

to a transcendence of personal goals in favor of communal cohesion. Social identity theory suggests that individuals derive their sense of self from membership in large, often dispersed social categories such as nationality, religion, and gender, shaping their interactions and perceptions within a community (Stryker et al., 2000). This process of identification reinforces in-group solidarity while distinguishing members from outsiders, strengthening the communal bonds that define collective engagement. Brewer and Silver (2000) expand on this idea, positing that participation in a shared identity leads to a cognitive shift from the individual "I" to the collective "we," where group identity becomes deeply embedded in self-concept and behavior.

In any social space, language functions as more than just a means of communication—it is also a mechanism of power and gatekeeping that influences participation and social hierarchy. The ability to understand and deploy dominant linguistic forms determines the extent to which users can navigate digital spaces, assert authority, and engage meaningfully with others (McDonald, 2015). The negotiation of language within these spaces reflects broader power dynamics, shaping not only who is seen as a legitimate member but also who holds influence over discussions and decision-making (Tufekci, 2017).

Social media platforms, including Reddit, amplify certain linguistic and cultural norms while sidelining others, reinforcing structures of exclusion and inclusion (Ananthasubramaniam et al., 2023). In these spaces, language serves both as a tool for fostering collective identity and as a barrier to participation. Users who are fluent in the dominant lexicons, inside jokes, and implicit rules of engagement within a subreddit are more likely to be recognized as valuable contributors and to hold informal influence within the group (Page, 2018). Newcomers who fail to align with these linguistic norms—whether through unfamiliarity with community-specific

terminology or misinterpretation of tone—may struggle to integrate, reinforcing an implicit hierarchy of knowledge and belonging (Gallagher & Savage, 2015).

The relationship between linguistic capital and digital power structures extends beyond individual interactions and into broader patterns of governance within online communities (Squirrell, 2019). At its core, the negotiation of language in digital spaces is a form of power—one that shapes access to information, authority, and social inclusion (Benson, 2009). The ability to define, enforce, and challenge dominant linguistic norms is essential for successful engagement in computer-mediated communication (CMC), where language serves as both a tool for interaction and a marker of identity. Unlike face-to-face interactions, digital spaces rely entirely on written language to establish community belonging and social cohesion. Mastery of platform-specific linguistic norms facilitates integration and recognition, while unfamiliarity can hinder participation (Danesi, 2014).

Social movements, political debates, and news dissemination are all influenced by who controls the language of engagement. Media institutions, platform moderators, and even influential users act as gatekeepers by determining what information is legitimized and what is discarded (Shoemaker & Vos, 2009). This is evident in the ways online communities mobilize around causes. Groups that successfully craft compelling narratives using dominant linguistics and symbolic frameworks gain visibility (Mendelsohn et al., 2024). Those that fail to do so struggle to attract engagement (Vu, 2014; Welbers & Opgenhaffen, 2018). In this sense, digital communities do not simply exist as neutral forums for discussion. They are arenas where power is negotiated, where access to dominant linguistic and cultural forms determines who has the ability to shape narratives and influence collective action (Barzilai-Nahon, 2008).

Bourdieu's (1991) concept of linguistic capital defines how fluency in dominant languages confers power in digital spaces. Just as proficiency in prestigious languages has historically provided social and economic advantages, control over the language of online discourse determines who can effectively advocate for their ideas and whose perceptions of culture are dismissed (Piller, 2016). This is apparent at the broadest level: the predominance of English as the language of global internet discourse disproportionately benefits those speakers with a high level of proficiency, reinforcing inequalities digitally (Phillipson, 2018). On a smaller scale, individuals and groups that communicate in marginalized languages face additional barriers in making their voices heard, reducing their ability to participate in global conversation and community-building efforts (Heller, 2011).

In their research on the punk scene in New Brunswick, New Jersey, researchers discovered that much of the punk community's reliance on secrecy was driven not by a genuine need to hide from institutions but by a desire to assert their opposition to those they considered outsiders, such as police officers who might threaten to shut down a house show (Lingel et al., 2012). The secrecy embedded into the language and communication practices within the notoriously secret and exclusive underground music scene was largely performative, used in combination with visual identifiers like clothing and hairstyles. The necessity of such alienation arose from the collective need to maintain a distinct group identity rather than any actual real threat to its existence.

The development of language and codes are not unique to underground subcultures. Phillips (2019) examines how trolling communities on 4chan and other platforms use irony and coded language as both a defense mechanism against external scrutiny and as a means of solidifying group identity. The use of "leetspeak" originated in online forums and gaming

communities as a form of symbolic resistance and identity formation (Stano, 2023). By substituting letters with numbers and symbols (e.g., “leet” becomes “1337”), members create an ingroup language that distinguishes them from outsiders and fosters a sense of belonging through demonstrated knowledge of culturally relevant indicators (Barasa, 2013). “Algospeak” is another, parallel concept used to describe the language used by social media users to circumvent moderation algorithms (Lorenz, 2022). By employing alternative spellings, symbols, or euphemisms, users are able to discuss sensitive topics without triggering automated censorship—maintaining community discourse and identity by relying on implied meanings (Klug et al., 2023; Steen et al., 2023).

The effectiveness of a community is often linked to the strength of its communication mechanisms. Research suggests that well-integrated communities rely on robust communication networks that enable members to engage in continuous dialogue, establish norms, and maintain a sense of shared purpose (Dijck et al., 2018). In contrast, fragmented or disorganized communities—those lacking cohesive communication structures rooted in shared identity—experience reduced engagement, lower social trust, and diminished capacity for collective action (Gil De Zúñiga et al., 2012). This distinction is particularly evident in digital spaces, where networked interactions determine the extent to which individuals participate and feel embedded within an online community.

Research on networked publics highlights the role of online discourse in fostering connectivity, suggesting that shared storytelling practices contribute to the development of community identity (boyd, 2010; Papacharissi, 2015). On platforms like Reddit, users engage with community narratives through intertextual discourse. They borrow and remix language, memes, and references to reinforce collective meaning (Milner, 2018). These discursive patterns

establish social cohesion while simultaneously delineating community boundaries, distinguishing insiders from outsiders based on their familiarity with shared linguistic and cultural cues.

However, the accessibility of information networks is not uniform across communities. Some online spaces provide well-defined pathways for participation, incorporating explicit norms and moderation structures that facilitate engagement, while others remain loosely organized, leading to fluctuating levels of interaction and visibility (Mercea, 2016). The ability of digital communities to sustain engagement over time depends not only on the availability of communication channels but also on the presence of mechanisms that regulate discourse, mitigate conflict, and reinforce social norms (Jenkins, 2013).

Information networks are central to the ways communities function, both online and offline. By providing the means through which members interact, share knowledge, and establish social norms, these networks sustain engagement and reinforce communal ties. The next section will explore how these dynamics intersect with platform governance and content moderation, examining the mechanisms through which online communities regulate discourse and maintain structural integrity.

Online Governance

Digital communities' ability to sustain engagement and cohesion relies on both communication and regulatory frameworks (Üblacker et al., 2024). Content moderation plays a critical role in maintaining these frameworks by mitigating behaviors that could undermine trust and participation (Ananthasubramaniam et al., 2023; Gruzd et al., 2023; Tobi, 2024). Effective moderation fosters a safe and inclusive environment, encouraging diverse voices to engage without fear while strengthening community bonds. Additionally, consistent enforcement of community guidelines enhances user trust and loyalty, as users are more likely to remain active

on platforms that prioritize their safety (Wilms et al., 2024). Conversely, platforms that lack adequate moderation often become breeding grounds for inflammatory and divisive content, eroding their ability to function as meaningful spaces for engagement (Stockinger et al., 2025).

Content moderation refers to the systematic process by which online platforms monitor, evaluate, and manage user-generated content to ensure compliance with established community standards, legal requirements, and ethical norms (S. Roberts, 2017). Over time, moderation has evolved from rudimentary community oversight into a sophisticated, multifaceted operation integral to modern digital platforms (S. Roberts, 2019). In the early internet era, moderation was largely voluntary, performed by community members who enforced basic guidelines to maintain civil discourse (Klonick, 2018). As digital spaces expanded, the scale and complexity of user-generated content increased, prompting platforms to develop formal content policies and employ dedicated moderation teams (Langvardt, 2018; Rozenshtein & Langvardt, 2025).

Fundamental understandings of digital content moderation

Digital communities, like physical ones, do not emerge in a vacuum. They require structures that facilitate coordination, establish norms, and maintain order (Boler, 2008). Unlike traditional communities, which form within geographic boundaries, online communities exist within digital infrastructures that must be intentionally designed, regulated, and maintained (Webber et al., 2022). These infrastructures take various forms, which Peters and Johnson (2016) categorize into three key types: (1) communication conduits that transport data across networks, (2) content hosts that store and provide access to third-party content (e.g., social media platforms like Reddit), and (3) search and application providers that organize and curate content without hosting it. The variation in these structures creates multiple approaches to content moderation,

making online governance a balancing act between free expression and accountability (Kidd, 2020).

Pre-moderation occurs when content is reviewed before it is made public, ensuring that posts align with platform guidelines before they reach an audience (Greis et al., 2014). This form of moderation is commonly associated with high-risk or highly regulated environments, such as news websites that require comments to be reviewed before publication or children's platforms with strict content controls (Myers West, 2018). Pre-moderation is particularly effective in preventing harmful or inappropriate content from surfacing, but it is also resource-intensive and can slow down the speed of communication, potentially discouraging engagement (Gongane et al., 2022). Pre-moderation mechanisms extend beyond direct content review. They can include structural barriers that filter users before they even participate. For instance, registration requirements, identity verification protocols, and CAPTCHA tests serve as pre-moderation tools by controlling who gains access to a platform and how they engage (Veglis, 2014). These mechanisms are often employed as deterrents to spam, bot interference, and inauthentic engagement, strengthening platform security while shaping the composition of digital communities (Gorwa et al., 2020). However, such restrictions also raise concerns about accessibility and privacy— particularly for marginalized users who may lack the required credentials or face risks associated with identity disclosure (Renaud & Coles-Kemp, 2022; S. Roberts, 2019).

Another key approach to pre-moderation is the use of automated filters and AI-driven screening tools to analyze content before publication. Platforms implement keyword filtering, sentiment analysis, and hash-matching techniques to detect potentially harmful language, hate speech, or misinformation before it is disseminated (Coutinho & José, 2017). While these tools

enhance efficiency, their reliance on algorithmic decision-making introduces risks such as false positives, contextual misunderstanding, and algorithmic bias (Davidson et al., 2017). For example, automated moderation has been found to disproportionately flag content from minority dialects and non-mainstream discourse communities, exacerbating concerns about over-policing and digital censorship (Sap et al., 2019). Algorithmic moderation extends beyond pre-moderation to regulate content at multiple stages of digital interaction. AI-driven moderation systems use machine learning to scan, classify, and sometimes remove content without direct human intervention. These systems operate at different levels, with some tools functioning before publication, while others work reactively in real-time by flagging and removing content post-publication. Platforms such as YouTube, Facebook, and Reddit employ predictive algorithms to detect violations of community guidelines and remove flagged content at scale (Jhaver et al., 2019)

The effectiveness of algorithmic moderation depends on the training data and governance models underpinning these AI systems. Bias in training data can lead to discriminatory enforcement, disproportionately penalizing certain linguistic groups, socio-political movements, or cultural expressions (Noble, 2018). Automated systems lack contextual awareness. They often struggle to differentiate between satirical content, reclaiming language in marginalized communities, and actual harmful speech. As a result, platforms frequently rely on a hybrid model that incorporates community feedback by combining AI-based moderation with human reviewers who provide contextual oversight (Udupa et al., 2023). While AI moderation enables scalability and efficiency, it remains an imperfect solution, as false positives and negatives continue to challenge enforcement consistency. Large-scale moderation failure, such as the inability of institutional level moderation to preemptively remove extremist content before it gains traction

on platforms, highlight the ongoing struggle to balance efficiency, fairness, and platform accountability (A. Marwick et al., 2022; G. Peters, 2022; Tech Against Terrorism, 2023).

Ultimately, pre-moderation and algorithmic moderation reflect competing priorities in digital governance. While they offer powerful tools for managing content at scale, their limitations underscore the need for greater transparency, adaptive human oversight, and participatory governance models to ensure that moderation systems align with user expectations and democratic principles (Suzor, 2019).

Content moderation serves a regulatory function—establishing guidelines, setting expectations, and adapting policies in response to community needs much like a governing body (Inserra, 2024). At the same time, moderators and automated enforcement mechanisms act as law enforcement entities, intervening in real-time to remove harmful content and sanction rule violations (Mazzurco, 2023). In this sense, digital platforms do not inherently embody democratic ideals but instead reflect varying degrees of governance, shaped by both platform policies and user participation (Forestal, 2022). The First Amendment of the U.S. Constitution protects individuals from government-imposed speech restrictions but does not extend to private entities, including social media platforms (Killion, 2024). This legal distinction places platforms in a unique position as both facilitators of speech and private regulators, responsible for setting and enforcing their own rules (Langvardt, 2018).

Scholars debate whether platforms should be considered neutral intermediaries or active participants in shaping public discourse. While proponents of platform neutrality argue that social media should function as open forums, critics highlight how content moderation decisions—such as algorithmic curation, deplatforming, and demonetization—can influence political discourse and economic opportunities (West et al., 2023). The economic incentives

behind moderation further complicate this debate. Research shows that platforms often balance content regulation with advertising revenue models. Stricter moderation policies may deter engagement but protect brand reputation by preventing association with harmful content (Liu et al., 2022).

The legal framework surrounding content moderation has been shaped by key court rulings that establish platform liability and moderation rights. One of the earliest cases was *Stratton Oakmont, Inc. v. Prodigy Services Co.*, in which Prodigy Services was held liable for defamatory statements posted by users because it exercised editorial control over its message boards (*Stratton Oakmont, Inc. V. Prodigy Servs*, 1995). This ruling distinguished platforms that actively moderate content from those that do not, setting a precedent for platform liability in online speech. The case underscored the legal complexities of content moderation, particularly regarding free speech concerns.

In response to growing concerns about platform liability, Section 230 of the Communications Decency Act became a pivotal moment in internet regulation (Communications Decency Act, 1996). This provision granted online platforms immunity from liability for user-generated content, allowing them to moderate content without fear of legal repercussions (Rozenstein & Langvardt, 2025). The passage of Section 230 enabled the enforcement of community guidelines, allowing platforms to balance free expression with harm mitigation. However, the law remains highly contentious, with ongoing debates over whether it enables platforms to evade responsibility for harmful content or whether it is essential for maintaining open digital spaces (Jiménez-Durán, 2023).

Historically, platforms have shouldered the responsibility of moderating content to maintain user-friendly environments (Boler, 2008). Dwork et al. (2023) propose a theoretical

model demonstrating that moderation policies not only encourage more diverse content but also foster healthier digital environments by mitigating harmful behaviors. Empirical studies further support this: Arttime et al. (2020) compared Twitter and Gab, finding that platforms with minimal or no moderation often descend into hubs for inflammatory discourse, weakening their capacity for constructive interaction.

This pattern is particularly evident in the rise of "alt-tech" platforms, such as Parler and Gab, which were created as ideological alternatives to mainstream social media. Gab, founded in 2016 by Andrew Torba, was explicitly positioned as a "free speech" platform, attracting a user base that included white nationalists, far-right extremists, and conspiracy theorists. Lax moderation policies allowed for the proliferation of hate speech, misinformation, and calls for violence, leading to its removal from major app stores and payment processors (Zannettou et al., 2018). Similarly, Parler—co-founded in 2018 by John Matze, with financial backing from Rebekah Mercer, a major funder of right-wing causes—branded itself as an alternative to Twitter with minimal content restrictions. It quickly became a hub for far-right organizing, particularly following the 2020 U.S. presidential election, when it played a significant role in facilitating extremist discourse leading up to the January 6th insurrection (Aliapoulios et al., 2021).

The lack of content moderation on these platforms not only enabled the spread of extremist content but also made them unsustainable in the long term. In the wake of the Capitol riot, Parler was banned from major hosting services, including Amazon Web Services, due to its failure to remove calls for violence (Nicas & Alba, 2021). While it later attempted a relaunch with revised moderation policies, its user base had already fragmented. Gab, on the other hand, remains operational but continues to struggle with mainstream acceptance and financial stability. These examples highlight how the absence of structured moderation frequently leads to the

proliferation of extreme content, eroding trust among users and making platforms increasingly volatile (Dowling, 2023).

Platforms often seek to minimize their role in content moderation, preferring instead to create an illusion of seamless user interaction (Gillespie, 2018). Direct intervention in speech regulation places them in a precarious position—too much moderation risks accusations of censorship, while too little invites legal scrutiny and reputational damage. At the same time, robust moderation requires significant investment in human labor and automated systems, which can be costly and difficult to scale (S. Roberts, 2019). To balance these competing pressures, platforms frequently rely on self-moderation strategies to effectively shift responsibility to users while still maintaining the broad legal protections under the Communications Decency Act (Potts et al., 2019). By framing moderation as a community-driven process rather than a platform-enforced policy, companies reinforce a model in which users—rather than platforms themselves—bear the burden of regulating digital spaces. This aligns with broader trends in the commodification of user activity, where engagement is not only monitored by monetization through data collection and targeted advertising (Gregorio & Goanta, 2023).

The current state of content moderation

Moderation labor has emerged as both a legal and ethical concern. Existing research highlights the psychological toll on human moderators who must review high volumes of distressing content (Arsht & Etcovitch, 2018). Content moderation also presents a technological challenge, as platforms face increasing pressure to develop automated moderation systems that can detect and filter problematic content at scale. However, algorithmic moderation introduces

new risks, such as bias, over-enforcement, and contextual misinterpretation (Cowgill & Tucker, 2019; Gorwa et al., 2020).

While automation reduces the burden on human moderators, it remains imperfect. AI-driven moderation often struggles with contextual nuance, leading to wrongful content removal or failure to detect harmful material. This problem is exacerbated by the proprietary nature of platform algorithms, which limits transparency in how moderation decisions are made (Chowdhury & Keller, 2022; Keller, 2022). As a result, users and policymakers continue to push for greater moderation transparency, ensuring that platforms provide clear guidelines, fair appeals processes, and explanations for moderation decisions (Lee, 2020).

Successful content moderation is dependent on context which, by nature, makes it specialized labor (Caplan, 2018). As such, human content moderators play a crucial role in maintaining the integrity of online platforms. Moderators filter out offensive or otherwise harmful content, dispute over decisions and recruit new moderators— all while constantly being asked to justify their decisions to the communities they oversee, the platforms they exist on, and their fellow moderators (Matias, 2019b). While some platforms do pay moderators for their labor it is ultimately a volunteer position— one that adds significant economic value (Postigo, 2003). Moderation can be manually demanding as well— clicking, scrolling, sorting through content takes time and effort (S. Roberts, 2016). Content moderation is not merely a set of rules for regulating behavior; it is a discursive practice that shapes broader narratives about identity, community governance, and platform authority. Moderation operates at multiple levels—legally, it determines the boundaries of permissible speech within private digital spaces (Article19, 2023). Economically, it influences engagement patterns and revenue structures (Rosenzweig, 2024). Socially, it plays a key role in defining community norms and reinforcing collective

identity (Team EMB, 2023). However, these functions are not always aligned, particularly as platforms increasingly rely on automated or large-scale moderation techniques. Currently, most algorithmic moderation efforts focus on formatting inconsistencies, spam detection, and copyright enforcement rather than addressing the complexities of harmful speech or cultural context (Gillespie, 2018).

The rapidly increasing scope and scale of social media membership poses one of the most significant challenges to in developing policy and practices at the to delegate responsibility for moderation practices. In February of 2025, California agreed to amend key provisions of a law that required large social media companies to disclose their policies on hate speech, disinformation, harassment, and extremism. This decision followed a challenge from Elon Musk on behalf of his platform, X, and a ruling by the 9th Circuit Court of Appeals that parts of the law likely violated the First Amendment (*X Corp. vs. Robert A. Bonta*, 2025). As a result, platforms are no longer mandated to report their definitions of hateful or misleading speech to state officials. They must, however, publish their terms of service and report changes biannually.

These are global challenges. In December of 2024, Vietnam implemented Decree 147 which imposes stringent regulations on social media platforms like Facebook, X, YouTube, and TikTok (Rising, 2025). Under Decree 147, users must verify their identity through phone numbers or ID cards and require platforms to suspend users and remove any content deemed illegal by Vietnamese authorities. Users may also make their own removal requests but are explicitly prevented from spreading any information about government wrongdoing. The Vietnamese government cites the need to limit the influence of content from foreign social media platforms, eliminate anonymous social media use, and increase responsibilities of social media platforms to moderate user engagement and content as reasoning for the passing of the decree.

Critics fear that policy such as this threatens free speech and provides impetus for mass surveillance of social media (Altman-Lupu & Swanton, 2025).

The United Kingdom has been actively addressing online misogyny and harassment through legislative and regulatory measures. In February 2025, Ofcom, the United Kingdom’s communications regulator, issued new guidance urging tech companies to enhance their efforts against “revenge porn” and explicit deep fakes. The guidance recommends the use of hash-hatching technology to identify and remove non-consensual intimate images in an attempt to curb the unauthorized sharing of such content (Davies & Milmo, 2025). Additionally, Ofcom emphasized the importance of implementing “nudges” to deter users from posting hurtful comments (Gill, 2023)— an approach very similar to X and Meta’s “Community Notes”. While Ofcom has no official capacity through which to enforce moderation practices, the United Kingdom has also passed legislation criminalizing the creation and distribution of sexually explicit deep fakes. The move addresses the significant rise in image-based abuse facilitated by artificial intelligence, often targeting women and girls. The new law enables prosecutions for both creating and sharing non-consensual explicit deep fakes via fines and imprisonment (Demony, 2025).

This development by the United Kingdom is one in a long string of attempts by governments to handle the onslaught of content created by artificial intelligence— whether for fear of misinformation or harassment. Several states in the United States have enacted similar laws, but there is currently no federal legislation prohibiting it (Graham, 2024). In September 2023, the “Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2023”, or DEEPFAKES Accountability Act, was introduced to the House of Representatives. The bill would protect national security against threats posed by

deepfake technology and provide legal recourse to victims of harmful deepfakes. The act proposed transparency requirements and established legal consequences for the creation and distribution of malicious deepfakes (DEEPFAKES Accountability Act, 2023). In 2024, the “Nurture Originals, Foster Art, and Entertainment Save Act,” or the NO FAKES Act, was introduced to the House of Representatives by Republican Congresswoman María Salazar alongside a bill of the same name to the Senate by Democratic Senator Chris Coons (NO FAKES Act, 2024a; NO FAKES Act, 2024b). This bipartisan legislation seeks to grant individuals a federal right to control digital replicas of their voice and likeness. Neither of these bills were passed, reflecting both an acknowledgment of necessity for governmental moderation protocol and hesitancy to interfere on such a large scale.

Moderators certainly play a crucial role as gatekeepers in their communities, but organizational factors are equally important in determining how and where information is shared (Stanoevska-Slabeva et al., 2012). These factors trickle down from the highest levels of federal government to individual actors. The interplay between structural governance and individual is particularly evident on Reddit.

Community, content moderation, and Reddit

While much of the existing research on digital communities conceptualizes Reddit as a social networking site, Hintz and Betts (2022) argue that this perspective is limiting. Reddit functions not just as a platform for connection but as a discursive space where digital interactions have tangible real-world consequences. The platform has drawn significant attention for high-profile incidents, such as the 2013 online witch-hunt for the Boston Marathon Bomber, which resulted in the wrongful identification of a suspect and their subsequent suicide (Hosek & Austin, 2016). Reddit also made headlines in 2021 for the GameStop stock short squeeze, during

which users in the subreddit r/WallStreetBets coordinated a mass purchase of GameStop stock. This caused extreme price volatility and forced major hedge funds—who had heavily shorted the stock—to suffer billions in losses, while many Reddit traders profited significantly (Anand & Pathak, 2021).

More recently, in June 2023, Reddit faced one of the largest coordinated user protests in the platform's history, known as the "Reddit Blackout." Thousands of subreddit moderators voluntarily shut down their communities in response to Reddit's decision to charge exorbitant fees for API access, effectively cutting off third-party applications and accessibility tools. The policy change disproportionately impacted visually impaired users who relied on third-party apps for screen reader compatibility, as well as moderators who used these tools for efficient subreddit management (Koebler, 2023). The blackout began as a temporary, two-day protest, but many large subreddits—including r/funny, r/gaming, and r/science—remained private for weeks in defiance of Reddit's corporate leadership. Some moderators sought long-term resistance by keeping their subreddits "dark" indefinitely, while others attempted more disruptive protest strategies such as changing subreddit themes or flooding their pages with unrelated content. Reddit administrators eventually removed moderators who refused to reopen their communities, asserting that subreddit ownership ultimately belonged to the platform, not the users (J. Peters & Sato, 2023).

The 2023 blackout underscored growing tensions between corporate platform governance and user-led community moderation. While Reddit has long marketed itself as a user-driven space where moderators enforce community-specific rules, the protest revealed the limitations of decentralized governance when corporate interests conflict with community autonomy (Schmitz & Samory, 2025). As platforms continue to evolve their monetization strategies, the event raises

important questions about the balance of power between corporate stakeholders and digital communities, particularly as moderation labor remains largely voluntary despite its critical role in content governance (Fiesler et al., 2018).

While high-profile events like the Boston Marathon bombing misidentification and the GameStop stock short squeeze highlight Reddit's capacity to influence real-world events, the platform's broader impact is deeply rooted in everyday, localized digital interactions. Within Reddit's ecosystem, each subreddit operates under its own set of community-created rules, fostering distinct cultures and norms that guide user behavior. This decentralized governance structure allows communities to tailor their guidelines to specific interests and values, thereby shaping the nature of interactions within them.

Reddit community culture

While platforms like Instagram, TikTok, and X prioritize personal profiles, follower networks, and algorithm-driven content recommendations, Reddit's default mode of interaction occurs within subreddits—user-created communities centered around specific interests, locations, or shared identities. This structural distinction places communities, rather than individual users, at the core of engagement, shaping how information is shared, debated, and moderated. Further, Reddit's platform encourages in-depth discussions over short-form content consumption (Mensah et al., 2020).

Subreddits are unique digital spaces that operate under a mix of centralized platform policies and decentralized, community-driven governance. The Redditors who make up these subreddits are a dedicated group of users that rely heavily on the site for a wide range of informational and social needs (Weld et al., 2024). The hybrid governance structure has made Reddit a valuable model for understanding how online communities self-regulate within larger

platform-imposed constraints. Users will leave subreddits according to their perceptions of the community and whether or not they feel their needs are being met (Hartzell et al., 2021). Similarly, communities change their rules as they grow and respond to new events (Reddy & Chandrasekharan, 2023). Further, the addition of new rules was found to be strongly associated with an immediate improvement in communities' perceptions of governance—signifying that moderators are consistently engaging with their community to better understand their needs and wants (Leibmann et al., 2025).

Written guidelines and visible moderation practices are crucial in shaping community within subreddits, but understanding community culture in online spaces also involves a recognition of the unwritten rules and social interactions that further guide such engagement (Centivany, 2016). These shared expectations define the overall vibe that each subreddit takes according to a number of different factors such as tone, topics of discussion, types of content, and even perceptions of and attitudes towards in-groups and out-groups (Oddný et al., 2023). However, the effectiveness of moderation policies is not uniform across communities—some subreddits employ stricter rule enforcement, while others prioritize flexibility to maintain user engagement (H. M. Wang et al., 2022). The balance between formal platform structure, moderation policies, and informal community norms ultimately shapes not only the quality of discussion but also the form that it takes.

Communication and information practices on Reddit

Discussions on Reddit unfold within subreddits, where shared interests, locations, and identities shape both the flow of discourse and the circulation of information (J. S. Park et al., 2022). The participatory structure of Reddit—where content visibility, rule enforcement, and moderation are determined by a blend of platform policies and decentralized community

governance—positions it as a critical site for examining how digital spaces function as self-regulated ecosystems. While many social media platforms rely on algorithmic curation, Reddit’s engagement model depends on participatory sorting mechanisms such as upvotes, downvotes, and user-moderated discussions. This process shapes not only what content is seen but also how knowledge is produced and validated in real time.

Reddit’s comment sections serve as primary spaces for interaction, often overshadowing the original post in importance. While initial submissions provide a foundation for discussion, the substantive exchange of information occurs within comment threads, where users add context, challenge perspectives, and refine the details of a topic (Straub-Cook, 2018). This comment-driven structure differentiates Reddit from more hierarchical platforms by fostering deliberative engagement, where collective information vetting occurs through layered conversation. Unlike traditional news sources, where editorial oversight determines credibility, Reddit communities rely on heuristic cues like karma scores, upvote ratios, and pinned moderator comments to signal reliability and shape the perceived legitimacy of shared content (Chandrasekharan et al., 2018).

These dynamics are especially visible in location-based subreddits, where communication practices intersect with information-sharing behaviors in ways that resemble both civic discourse and localized news distribution (Riley & Cowart, 2021). Users turn to these communities for a wide range of inquiries, from routine requests like restaurant recommendations or housing advice to real-time updates on public safety incidents, transportation disruptions, and government decisions. This mirrors the informal social networks that have historically played a role in local knowledge dissemination, where trust is built through repeated interactions and shared

community experiences rather than through official verification mechanisms (Amini et al., 2024).

Participation in these communities is contingent upon both explicit and implicit norms that regulate discourse and establish credibility. Unlike platforms that rely primarily on personal reputation, Reddit fosters a system where adherence to subreddit-specific norms dictates user influence (Deutsch & Gerard, 1955; Hogg, 2010). User behavior like comment frequency, voting history, and engagement with community rules shapes visibility and authority within a subreddit, reinforcing self-regulatory mechanisms that determine who is seen as a legitimate contributor. This is particularly evident in location-based subreddits, where insiders with deep local knowledge often serve as *de facto* information gatekeepers, signaling belonging through references to local landmarks, slang, and historical context. Newcomers or outsiders who fail to align with these implicit expectations may struggle to gain traction, as their contributions are met with skepticism, downvotes, or moderator intervention (Danescu-Niculescu-Mizil et al., 2013).

While governance structures on Reddit broadly influence user behavior, location-based subreddits offer a distinct case study in how digital communities balance communication and information practices. Unlike interest-based subreddits, which engage in broad discussions that may not be rooted in immediate lived experiences, location-based subreddits serve as digital extensions of physical communities. As local newsrooms shrink and traditional journalism recedes, residents increasingly turn to these platforms to negotiate what is locally relevant. Nikki Usher (2021) argues that the decline of local journalism has fractured civic engagement, pushing communities toward nationalized narratives while leaving gaps in local discourse. Without dedicated local reporting, location-based subreddits function as alternative information networks, where news is not just consumed but actively produced and interpreted by the community itself.

The intersection of communication and information practices in these spaces reveals the evolving nature of digital civic engagement. Moderation decisions, whether enacted through explicit rule enforcement or informal community consensus, shape what information circulates and how it is framed. The ability of subreddit communities to self-regulate through upvotes, discussion norms, and collective content evaluation provides a compelling alternative to traditional editorial gatekeeping. Atton & Hamilton (2008) argue that community-driven media, particularly those reliant on non-professional actors, prioritize values such as place, tradition, and social cohesion over ideological divides. While digital communities allow for the transcendence of geographic limits, locality is still an important part of informing communication practices that inform individual and community identity.

By examining how location-based subreddits facilitate both discourse and information-sharing, this study contributes to a broader understanding of how digital platforms shape community identity, governance, and participatory engagement. As major social media companies shift toward less oversight in the name of "free expression," Reddit remains an alternative model—one where decentralized governance has proven functionally effective in maintaining order while fostering community cohesion. Understanding how digital spaces self-regulate, particularly in location-based contexts, offers insight into the broader role of participatory governance, digital identity, and collective engagement in online civic discourse.

Summary and next steps

This study focuses on Reddit as a case study, investigating its cultural norms, moderation practices, and the discursive nature of participation within subreddits. Unlike other social media platforms that prioritize personal profiles and algorithm-driven content curation, Reddit's structure places communities at the center of engagement. The ways in which subreddits manage

discourse, enforce norms, and foster identity formation provide a compelling lens for understanding digital civic participation. This is particularly relevant in location-based subreddits, where online governance mirrors traditional patterns of local media and community organization.

This literature review examined the evolution of digital communities, the role of content moderation in shaping online spaces, and the governance structures that define Reddit. From theories of community formation to modern frameworks of digital governance, research shows that online communities are not just spaces for interaction but self-regulated ecosystems where norms, rules, and identities are constantly negotiated. These insights inform the present research, which explores how content moderation shapes discourse on location-based subreddits as democratic communication forums. This study seeks to answer the following questions:

1. How do moderation practices in location-based subreddits impact the discourse that occurs within them?
2. How do moderation practices and guidelines shape community identity in location-based subreddits?
3. To what extent does community identity inform moderation practices in location-based subreddits?

By analyzing discourse patterns, moderation structures, and engagement behaviors, this research contributes to ongoing discussions about digital governance, platform accountability, and online civic engagement. The following methods section outlines the research design, data collection process, and analytical approaches that will guide this study.

CHAPTER 3

METHODOLOGY

This study applies a thematic content analysis to examine how moderation practices and community identity interact within location-based subreddits. Prior research highlights the negotiated nature of platform governance, where moderation structures both constrain and enable discourse, shaping how users participate in digital spaces. By analyzing patterns of engagement, rule enforcement, and community norms, this study explores how these factors influence the types of discussions that emerge and the ways subreddit members construct their local identities.

Through the systematic coding of posts across four subreddits, this approach provides insights into how users interpret and respond to moderation, what types of content generate the most discussion, and how subreddit culture influences engagement. These findings contribute to broader discussions about the relationship between digital governance and participatory culture, particularly in decentralized online spaces where users play a direct role in shaping discourse. In the following sections, I outline the data selection, collection, and coding strategies used to analyze how moderation structures shape community identity and how subreddit members navigate and contest these rules in practice.

Thematic content analysis

Thematic content analysis (TCA) is a qualitative research method used to identify, analyze, and interpret patterns within textual data (Braun et al., 2019). It provides a flexible yet systematic approach to examining qualitative data by categorizing themes that emerge from the

dataset. TCA several key steps: (1) familiarization with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report (Braun & Clarke, 2006). This iterative process ensures that themes are not just identified based on surface-level patterns but also contextualized within broader social and cultural discourses. Thematic analysis is particularly useful in digital ethnography and online discourse studies, where patterns of engagement and community norms emerge through user interactions (Markham, 2021). In the context of this study, TCA is used to examine how moderation practices and community identity intersect within location-based subreddits. By systematically coding and categorizing posts, this method enables the identification of recurring themes related to governance, information-sharing, and digital community norms. Following an inductive approach, themes were derived directly from the data rather than being imposed by predefined categories. This approach aligns with digital ethnographic practices, which emphasize immersion in online spaces to understand how users collectively construct meaning and enforce norms (Boellstorff, 2012). Nowell et al. (Nowell et al., 2017) argue that TCA is especially valuable in studies of digital culture because it allows researchers to make sense of vast amounts of online data while retaining contextual nuance. However, as Vaismoradi et al. (2013) caution, thematic analysis requires rigorous attention to coding reliability, reflexivity, and transparency to ensure that findings remain credible and representative of the dataset.

By employing digital ethnographic study rooted in thematic content analysis, this study captures the implicit and explicit mechanisms of community moderation, shedding light on how users engage with and negotiate subreddit guidelines. This method allows for an in-depth exploration of the ways in which digital communities self-regulate, interpret rules, and establish

collective identity, contributing to broader discussions on participatory governance and online moderation.

The case for digital ethnography

Changing spaces means changing the way we interact with and interpret them. As such, reconceptualizing traditional methods is crucial for effective research. This study combines traditional qualitative methods with digital ethnography, a research approach that emphasizes sustained engagement in online communities to understand group dynamics, cultural norms, and knowledge production (Hine, 2015; Pink, 2016). Digital ethnography allows researchers to study networked interactions, platform-specific behaviors, and how digital spaces function as social and informational hubs (Caliandro, 2018). Scholars like Nardi (Nardi, 2016) and Lingel (Lane & Lingel, 2022; Lingel, 2021) argue that embedding oneself within online communities is essential for capturing the social and communicative practices unique to digital environments, particularly in decentralized and community-driven spaces like Reddit.

Much like traditional ethnographers immerse themselves in physical communities, digital ethnographers observe, analyze, and sometimes participate in online interactions to understand how individuals negotiate community identity (Nardi, 2010). In her study of digital countercultures, Lingel (Lingel, 2017) demonstrated how persistent engagement fosters credibility and deeper insights into highly insular online subcultures, including body modification, punk, and drag communities. By maintaining a visible presence in relevant chat rooms and public forums, she gained access to insider perspectives on content-sharing practices and community self-governance. Hine (2015) similarly argues that digital ethnography provides a critical lens for understanding platform affordances, community expectations, and user

behavior in context, rather than treating digital content as isolated from the social structures that produce it.

In the context of this study, digital ethnography was used to observe how moderation and information-sharing operate within four location-based subreddits. Approximately three months before data collection, I subscribed to each subreddit from an account created specifically for this study, enabling notifications for new activity and regularly checking them to stay apprised of community happenings. While I did not participate in discussions, this passive immersion allowed me to experience the subreddit as a regular member—observing content trends, moderation decisions, and community norms as they unfolded in real time to inform the creation of codes and ultimately provide further context within which to orient my findings. This approach aligns with Pink’s (2016) argument that digital ethnography requires ongoing engagement rather than a single-point analysis, ensuring that research captures both the structural and dynamic elements of online communities.

This method also acknowledges Nardi’s (2010) emphasis on understanding platform affordances and technical infrastructure as central to digital ethnography. Reddit’s algorithmic ranking system, moderation tools, and community governance structures all influence discourse and community identity. By embedding myself within these spaces before conducting formal analysis, I was able to contextualize content patterns within the broader framework of platform structures and subreddit culture.

Data Selection

When selecting cities for this study, multiple characteristics were considered to ensure a balanced and comparative analysis of location-based subreddits. Each subreddit needed similar membership sizes to maintain comparability in data collection. The selected subreddits all had

between 470,000 and 570,000 members, ensuring that observed differences were due to community identity and engagement patterns rather than variations in subreddit size. Because Reddit user data is anonymous, city-level demographic markers—such as geographic region, racial composition, median income, education level, and political affiliation—were used as proxies for understanding potential differences in subreddit culture. U.S. Census data was collected and analyzed to contextualize the digital engagement patterns observed in each subreddit.

Table 1. *Data selection metrics*

City	Population	Median Income	Education (bachelor's or higher)	Employment	Age	Racial majority	/subreddit users
Austin	961,855	\$ 91,505.00	62%	72.60%	34.9	White	473,00
Chicago	2,746,288	\$ 74,474.00	46%	45.70%	36.4	White	571,00
Philadelphia	1,603,797	\$ 60,302.00	36%	60.20%	35.6	Black	517,00
San Francisco	873,965	\$ 126,730.00	60%	65.90%	40.7	White	521,00

The study aimed to incorporate cities that reflected a range of regional, economic, and social contexts rather than serving as a perfectly representative sample of the U.S. population. While cities tend to share some characteristics, like higher-than-average education levels and urban political leanings, each selected city embodies distinct cultural and socioeconomic dynamics that contribute to differences in digital discourse. Philadelphia, on the East Coast, has a predominantly working-class population with a strong tradition of local political engagement. San Francisco, on the West Coast, is characterized by its affluent and highly educated demographic, shaped by its role as a tech hub. Chicago, a Midwestern metropolis, provides both geographic and demographic balance— capturing elements of both coasts while maintaining its own urban identity. Austin, a fast-growing Southwestern city, adds further cultural and geographic diversity to the sample.

Balancing comparability with demographic and regional diversity was a key methodological challenge. While the study does not attempt to create a fully representative cross-section of the U.S. population, it does aim to explore how local contexts shape digital engagement. The selected cities span different regions, economic compositions, and urban cultures, allowing for a nuanced analysis of community identity formation in online spaces. The decision to focus on cities rather than states, neighborhoods, or entire nations was deliberate. Cities provide a manageable unit of analysis, where demographic trends are more consistent, and discussions of local issues are clearly defined. State-level subreddits introduce excessive political and cultural variability, while neighborhood-based subreddits often have too few active users to allow for meaningful comparative study. City-based subreddits strike a balance, offering a broad enough scope to analyze patterns of engagement while still reflecting localized discourse and governance and clearly defined geographic boundaries.

The data collection period was selected to capture a broad range of community interactions while ensuring the findings were both representative and analytically useful. Posts were collected in two waves, first on December 1, 2024, at 5 p.m. and again on December 20, 2024, at 5 p.m., with each pull retrieving the most recent 1,000 posts per subreddit. This two-wave collection method was chosen to observe patterns of engagement over time rather than relying on a single snapshot, which could have disproportionately reflected short-term anomalies. By spanning two distinct points in the calendar, the study was able to assess whether trends in discourse remained consistent or shifted due to external events, seasonal factors, or changes in moderation enforcement.

One key consideration in selecting this timeframe was the immediate aftermath of the 2024 U.S. election cycle. Elections frequently spark discussions in local forums, particularly

regarding governance, civic participation, and the broader political climate. Observing subreddit activity in the weeks following an election provided an opportunity to analyze how community guidelines shaped political discourse, particularly in spaces where discussions of national politics were often restricted in favor of locally relevant topics. This period allowed for the examination of moderation enforcement related to political content, as well as user responses to those enforcement mechanisms, revealing the extent to which political discussions were shaped by both explicit subreddit rules and informal community norms.

Beyond its political significance, the chosen data range also captured seasonal patterns in community engagement. Late November through mid-December represents a period of heightened social and economic activity, as users seek recommendations for holiday events, shopping, and travel. This time of year generates a distinct surge in location-based inquiries, ranging from requests for holiday-specific services, such as where to purchase Christmas trees or find seasonal markets, to broader discussions of local traditions and celebrations. The presence of these posts in all four subreddits demonstrated how cultural engagement and community identity were reinforced through user participation in annual traditions.

By carefully selecting this data range, the study was able to capture a diverse and representative cross-section of community activity within location-based subreddits. The combination of political discourse, seasonal engagement, and everyday information-sharing allowed for a comprehensive examination of the ways in which moderation practices and community identity intersect in online spaces. Through this selection, the study ensures that the findings are not only reflective of momentary trends but also provide insight into broader patterns of user behavior within digital localities.

The two-wave collection method was also designed to mitigate limitations posed by Reddit's API, which restricts data pulls to the most recent 1,000 posts at the time of collection. This means that the temporal range of each dataset varied depending on the overall activity level of the subreddit. A more active subreddit would reach the 1,000-post limit more quickly, meaning its dataset would cover a shorter period, while a less active subreddit would have data stretching back further in time. Conducting two separate pulls allowed for a more balanced representation of subreddit activity, ensuring that engagement trends were not disproportionately shaped by fluctuations in posting frequency on any given day.

Data Collection

The process of data collection for this study was designed to ensure consistency, comparability, and methodological rigor. Given the study's focus on community moderation and digital governance within location-based subreddits, it was essential to employ a systematic approach to retrieving posts while adhering to platform constraints. To accomplish this, Reddit's open-source API (Application Programming Interface) was used to extract posts directly from the four selected subreddits: r/Austin, r/Chicago, r/Philadelphia, and r/SanFrancisco. By leveraging the API, the study was able to collect real-time data while maintaining the integrity of the original content, free from the potential biases introduced by external aggregation tools.

Python 3.10 was used to execute the script that interacted with Reddit's API, with the Python Reddit API Wrapper (PRAW) serving as the primary tool for authentication and data extraction. Each API request was configured to return the most recent 1,000 posts from each subreddit, in accordance with Reddit's data retrieval limits. Once the posts were retrieved, they were stored in a structured format using the Pandas library, which facilitated subsequent cleaning

and analysis. Each post's metadata—including title, content, timestamp, engagement metrics (upvotes, downvotes, comments), and subreddit-specific features such as flair—was preserved in a CSV file for further coding and interpretation. The collected data was then reviewed to ensure completeness and accuracy, with redundant or incomplete entries removed as necessary.

While the study's methodology prioritized the collection of complete datasets, some limitations inherent to Reddit's API had to be addressed. The most significant restriction was the API's inability to retrieve posts that had been deleted or removed before the time of collection. This meant that content flagged as rule violations by subreddit moderators or removed by Reddit's automated moderation systems may not have been captured. Although the study attempted to account for deleted content by identifying potential violations through user discussions and moderator responses, it was not possible to retrieve all missing posts. This limitation underscores the challenge of studying moderation practices in real-time digital environments, where content removal occurs rapidly and without permanent archival mechanisms.

In total, the two rounds of data collection resulted in a dataset of 7,603 posts across the four subreddits. The specific distribution of posts varied by subreddit due to differences in activity levels. More active subreddits reached the 1,000-post limit more quickly, meaning their datasets covered a shorter time span. Conversely, less active subreddits had posts stretching further back in time. This variability in dataset coverage was taken into account when analyzing engagement patterns, ensuring that conclusions were drawn with appropriate contextual awareness.

Table 2: *Total size and date range of dataset*

Subreddits	Number of posts collected	Date
Austin	1,959	11/12/2024-12/20/2024
Chicago	1,835	10/20/2024-12/20/2024
Philadelphia	1,880	10/25/2024-12/20/2024
San Francisco	1,929	11/13/2024-12/20/2024

Another key factor in the data collection process was the inclusion of post-specific metadata, which played a crucial role in understanding moderation dynamics. The script pulled for the title of the post, the body of the post, the date posted, and a link to the original post within the subreddit. Additionally, engagement metrics such as vote and comment counts provided further context for understanding how posts resonated within each community. By utilizing Reddit's API and implementing a methodologically rigorous retrieval process, the study captured a rich dataset that reflects the complexities of digital civic engagement. This dataset forms the foundation for the thematic analysis that follows, offering a detailed examination of how moderation practices shape discourse within online communities.

Coding

After the posts were collected, they were coded using a structured framework informed by both existing literature and an inductive approach to thematic content analysis. The coding process aimed to identify patterns in user interactions, content types, and community moderation practices across the four selected subreddits.

Development and application

The method followed a grounded theory approach, commonly employed in qualitative social research to develop conceptual understandings directly from observed data (Birks & Mills, 2015; Glaser, 1965). This approach allowed for flexibility in refining the coding structure as patterns emerged, ensuring that the framework remained responsive to the complexities of online community discourse. Rather than imposing rigid predefined categories, the coding process evolved through iterative engagement with the dataset, aligning with best practices in thematic content analysis (Braun & Clarke, 2006). While thematic analysis has been widely used in media and communication research, scholars have debated its conceptual boundaries, given that it often functions as an umbrella term encompassing multiple qualitative methodologies (Nowell et al., 2017). In this study, thematic content analysis served as a tool for identifying patterns in user interactions, rule enforcement, and community engagement, providing a structured yet adaptable lens for examining how digital communities shape discourse.

To develop the coding framework, an initial pre-test was conducted on a sample of 400 posts from r/Austin. This preliminary analysis provided insights into the most frequently recurring discussion topics and engagement patterns, informing the creation of an initial set of codes. These codes were refined through inductive coding, a process in which themes emerged organically from the dataset rather than being strictly predefined. Each post was classified according to both content and function, ensuring that the coding process captured not only what users were discussing but also how they engaged with their communities. Posts were analyzed not in isolation but in the broader context of subreddit norms, moderation trends, and interactions between users.

Table 3: *Classification of codes*

Category	Content	Function
Information	• Recommendations • "What's that?" posts • moving/visiting	Request information
Public interest	• Found pets • Politics • Growth and development	Share information
Culture	• Observations • Jokes	Neutral

A post was labeled as information when the primary goal of the user was to seek specific knowledge, guidance, or recommendations. Many of these posts asked about services, events, or logistical concerns such as moving. Questions about public transportation, weather conditions, or recommendations for local businesses were common. Public interest posts sought to share information with the wider community. These posts frequently shared local news, safety alerts, or discussions about civic policies. They addressed issues of governance, crime reports, public service updates, or social movements within the city. Culture posts centered on identity, shared traditions, and local humor. They were often expressive rather than transactional, fostering a sense of belonging among users. Some posts highlighted city-specific quirks, whether through inside jokes, long-standing traditions, or humorous observations about daily life. Others reflected nostalgia, as users reminisced about past versions of their city and how it had changed over time.

Some posts did not fit neatly into a single category and required dual coding. A post that included a joke about the city's infrastructure, for example, might be categorized as both cultural and public interest. Thematic coding in these cases prioritized the intent of the user rather than just the structure of the post. A meme about rising rent prices in San Francisco might serve as

both cultural commentary and a critique of urban policy, meaning that it functioned simultaneously as humor and as civic discourse. The fluidity of these discussions highlights the difficulty in imposing rigid classifications on online communities, where meaning is often negotiated in real time through user interactions.

Certain posts were also coded based on their level of engagement rather than just their thematic content. High-engagement posts stood out as anomalies, drawing significantly more attention than the average post within a given subreddit. These posts often amassed high upvote counts, extensive comment threads, or frequent cross-posting to other subreddits. The reasons for such engagement varied. Sometimes, the appeal stemmed from the content itself. An especially compelling or provocative post was likely to generate discussion. Other times, the subreddit's collective interests or frustrations shaped the level of participation. The coding process paid close attention to these high-engagement posts, as they often provided insight into what resonated most with the community. Due to difference in subreddit size, high engagement was qualified relative to trends present within each subreddit.

Moderation played a crucial role in shaping the discourse within these subreddits, making it important to track posts that appeared to violate community rules. These posts were tagged as potential rule violations when they contained potentially sensitive content. Common violations included off-topic content, political posts outside designated discussion threads, and self-promotion, particularly when users attempted to advertise businesses or fundraisers. Posts that contained inflammatory language, harassment, or misinformation were also coded for potential violations, though it was not always clear whether they had been removed due to direct moderation or through downvotes by the community.

Tracking rule violations was complicated by Reddit's reliance on automated moderation tools. The AutoModerator system removed some posts before they had a chance to be widely seen, making it difficult to analyze the rationale behind removals. Additionally, because Reddit's API does not provide access to deleted posts, it was impossible to retrieve content that had been removed prior to data collection. This limitation meant that some patterns in moderation enforcement had to be inferred through user discussions about deleted content rather than through direct observation.

After the initial round of coding, a secondary review was conducted to refine ambiguous classifications. Posts that could be categorized under multiple themes were reviewed to ensure consistency. This secondary review was particularly useful in cases where discussions evolved over time, shifting in focus as users engaged with the content. A post that began as an information request might, through an extended comment thread, turn into a broader cultural discussion. By revisiting these posts, the coding framework was able to better capture the full spectrum of interactions within each subreddit.

Despite the limitations posed by Reddit's moderation system and data access restrictions, the coding process provided a detailed understanding of how location-based subreddits function as semi-regulated spaces for local discourse. It revealed patterns in how users sought information, engaged with civic topics, and constructed cultural identity through digital interaction. The findings from this analysis offer insights into how community norms shape discussion, how moderation influences engagement, and how users collectively negotiate the boundaries of acceptable discourse in digital spaces.

Summary of methods

This methodological framework provides a structured yet adaptive approach to understanding how moderation practices and community identity intersect within location-based subreddits. By applying thematic content analysis alongside digital ethnographic methods, this study examines the complex interplay between governance, user engagement, and collective identity in online spaces. The iterative coding process allowed for the identification of recurring themes related to information sharing, civic discourse, cultural participation, and moderation enforcement, offering insight into how subreddit communities regulate themselves and negotiate their own norms.

The research questions guiding this study emphasize the relationship between moderation practices and community identity, the role of subreddit guidelines in shaping discourse, and the extent to which community norms influence moderation decisions. The findings presented in the next section demonstrates how subreddit members actively contribute to the governance of their digital spaces. By exploring patterns of engagement, rule enforcement, and collective identity formation, the findings section will offer a comprehensive examination of how location-based subreddits function as dynamic arenas of digital civic participation.

CHAPTER 4

FINDINGS

This section presents the findings of this study, examining how content moderation and community identity interact within and across four location-based subreddits: r/SanFrancisco, r/Austin, r/Philadelphia, and r/Chicago. Content was categorized into three primary types: information requests, public interest discussions, and cultural discourse. Additional codes were applied to posts that either potentially violated subreddit rules or generated unusually high engagement, defined as posts that received significantly more upvotes, comments, or visibility than the subreddit's average within the selected timeframe.

The findings reveal that local identity and cultural priorities shape both the types of content shared and the ways discussions unfold at a substructural level through the shared enforcement and negotiation of community identity. Each subreddit constructs its own regulatory framework, defining acceptable discourse and establishing rules that reproduce the values of its members. By situating engagement patterns within the specific moderation cultures and rules governing each community, this section explores the intersection of digital governance, local culture, and participatory discourse.

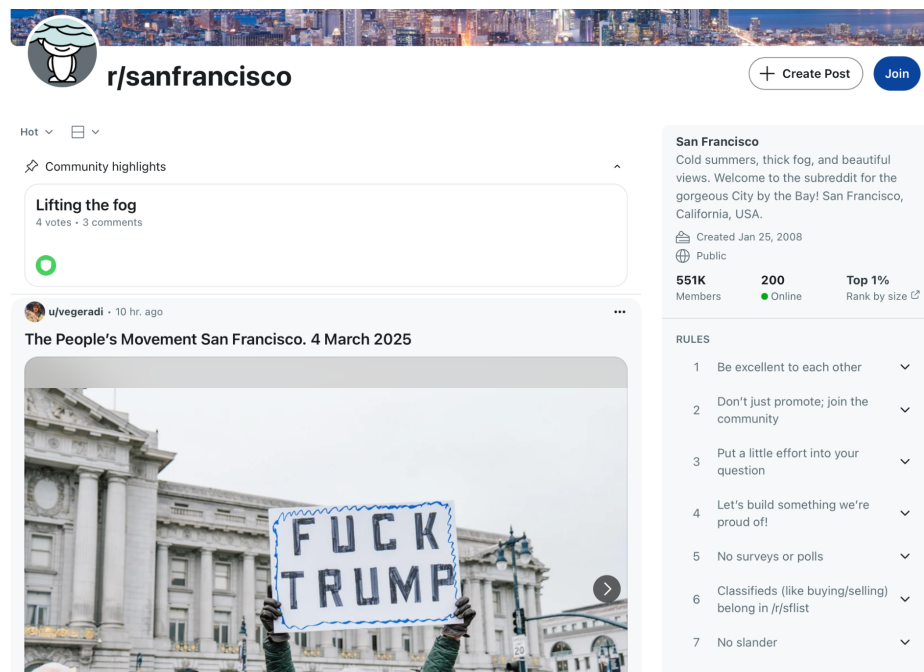
Each subreddit is moderated by volunteer users, not Reddit employees, who set and enforce rules based on subreddit-specific priorities. Moderators can issue warnings, remove posts, and ban users. Additionally, two key platform features—karma and flair—help to further structure user participation and content organization. Karma serves as a numerical representation of a user's contributions to the platform. Users earn post karma when their submissions receive

upvotes and comment karma when their responses are upvoted. Conversely, downvotes decrease karma, signaling that a contribution is either unhelpful, irrelevant, or against community norms. While karma has no official function beyond reputation-building, some subreddits impose minimum karma requirements to limit participation, preventing brand-new or low-reputation users from posting or commenting. In some cases, moderators use negative karma as a warning sign, scrutinizing or restricting users with a history of unconstructive engagement.

Flair is a labeling system that helps categorize content and identify users. Subreddit-specific post flair allows moderators or users to tag submissions with predefined labels, such as “Discussion”, “News”, or “Help”. This makes it easier to navigate subreddit content. Some subreddits also use user flair, which appears next to usernames and can reflect location, expertise, or other identifying details relevant to the community. While some flairs are manually assigned, others are automated through AutoModerator. AutoModerator is a bot that enforces subreddit rules, detects violations, and applies labels based on keywords or content type. Together, karma and flair shape subreddit dynamics by influencing visibility, credibility, and content organization. While these features are incorporated at the platform level, it is up to moderators to determine how they will apply within individual subreddits and communicate that effectively to subreddit members.

The consequences for rule violations vary, with some infractions leading to stricter penalties than others. Repeat offenders may face harsher actions, including longer bans or permanent removal from the subreddit. Each subreddit has a guidelines section, found in the “About” tab on the right side of the site’s layout.

Figure 1. *Subreddit layout*



While the structure and specificity of these rules vary, most subreddits organize their guidelines into clearly defined categories addressing key aspects of community engagement. Larger subreddits often supplement their guidelines with frequently asked question (FAQ) sections or sticky posts to address common concerns. Other common practices include providing links to external Reddit pages that provide further information. The “About” tab also includes a list of the moderators with an option to message them individually or as a team, however moderation teams may opt to hide this list. Within this findings section, “Guidelines” refers to the list of rules as a whole. “Rules” refer to specific points within these lists.

While Reddit moderators are not employees and Reddit as a platform does not dictate subreddit guidelines in any official capacity, it does promote a user-drive approach to platform etiquette through a widely acknowledged set of guidelines known as “Reddiquette”. Reddiquette is a community-driven code of conduct that outlines best practices for engagement. It is

described by Reddit as “an informal expression of the values of many redditors, as written by redditors themselves” (*Reddiquette*, 2025). A core tenet is the principle of “Remember the human”, which encourages users to communicate online as they would in face-to-face interactions. Users are expected to adhere to the same behavioral standards online as in real life, engaging in good faith and avoiding personal attacks, harassment, or disruptive behavior. Before contributing, users should read subreddit rules and familiarize themselves with community expectations as these vary widely across Reddit.

Voting is central to Reddit’s functionality, and *Reddiquette* advises users to vote based on content quality rather than opinion. This ensures that well-written and meaningful contributions are not downvoted simply due to disagreement. Additionally, submission titles should be factual and free of editorial bias, and users are encouraged to seek out and link to original sources rather than secondary aggregators. To avoid redundancy, users should search for duplicate posts before submitting content and post to the most appropriate subreddit for their topic.

Participation on Reddit should be transparent and constructive, according to *Reddiquette*. Self-promotion is allowed within reason, following the widely accepted 9:1 ratio. Only one out of ten posts should be self-promotional. When downvoting content, users are encouraged to provide constructive criticism rather than simply dismissing posts without explanation. Users should also report spam and rule-breaking content rather than engaging in flame wars or vigilante moderation. Certain behaviors are strictly discouraged. *Reddiquette* explicitly forbids posting personal information (doxxing), engaging in illegal activity, harassing other users, trolling, or inciting mass voting campaigns. *Reddiquette* also addresses NSFW (Not Safe for Work) content. This refers to posts containing explicit material, including nudity, violence, or

graphic injury. While NSFW content is allowed in many subreddits, failing to mark it properly can result in removal or penalties. Some subreddits prohibit NSFW material altogether.

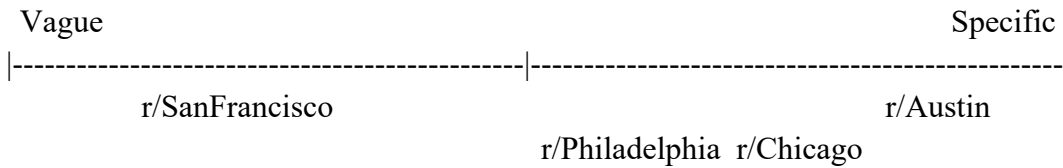
Additionally, users should refrain from reposting deleted or removed content, as these have typically been taken down for valid reasons. Moderators are expected to enforce rules fairly and should avoid moderating in communities where they have conflicts of interest. The guidelines also extend to commenting practices, discouraging low-effort remarks such as “this” or “lol” and urging users to provide meaningful contributions. Complaining about votes, reposts, or posting across subreddits is also frowned upon, as Reddit’s ranking system naturally determines what content gains visibility. Users are reminded to edit their posts transparently by indicating changes with an “Edit:” tag to maintain clarity in discussions.

Reddiquette is an evolving document that adapts as Reddit’s communities grow and change. While adherence is voluntary, it provides a framework for maintaining respectful, engaging, and well-moderated discussions, helping users navigate the platform’s social norms and expectations.

Community guidelines

The four subreddits studied operated within Reddit’s broader platform regulations but differed in how specific their moderation guidelines were. The scale below reflects these differences based on the number of rules each subreddit had and how detailed their descriptions were. Some outlined broad principles with room for interpretation, while others set clear rules users were expected to follow. Figure 2 illustrates this variation—positioning each community along a spectrum from vague to specific. r/SanFrancisco had the fewest and least detailed rules, relying on broad principles. r/Philadelphia and r/Chicago, fell in between, offering a mix of general expectations and specific enforcement policies.

Figure 2: *Guideline classification*



Each subreddit articulated its own unique expectations, reflecting a distinct approach to defining community norms and behavioral standards. Subreddit guidelines touched on a wide variety of topics. A general breakdown of them is below, the topics listed are based on content included in the platform-wide Reddiquette. Moderation guidelines made mention of a wide variety of topics— some overlapped and some were community-specific (Table 4).

Table 4: *Content of subreddit guidelines*

	Austin	Philadelphia	Chicago	San Francisco
Contact with moderators	x			
Personal information	x			
FAQs	x	x	x	x
Community relevance	x	x	x	x
Self-promotion	x	x	x	x
NSFW		x	x	
Editorializaton	x	x	x	
Harassment	x	x	x	
Provision of sources	x	x	x	
"Be Nice"			x	
Homelessness	x	x	x	
Crime	x	x	x	x

Subreddit guidelines vary in length and depth. Some had shorter lists of rules but more detailed descriptions of each one. Some include links to other interfaces with further information. These differences were also included when considering which subreddits were deemed vague to specific.

r/SanFrancisco

r/SanFrancisco had the shortest guidelines, with only seven rules total. It encourages its users to “Build something we’re all proud of”. The description following this rule (number four in the subreddit’s list of seven guidelines), provides a more detailed description of what is meant by this. Moderators tell users that they can make San Francisco a better place, specifically, by using the subreddit to volunteer and organize, posting San Francisco-themed artwork, organizing Zoom meetings with fellow users, contacting moderators with ideas, or to “gather a crowd from some bizarre stunt” (*San Francisco*, n.d.). While the general sentiment is vague, a clear image of what San Francisco is (and is not) exists just below the surface. Additionally, the “About Community” section in San Francisco simply states “Cold summers, thick fog, and beautiful views” as a description of the subreddit (*San Francisco*, n.d.) .

Other notable guidelines from r/SanFrancisco include rule number one: “Be excellent to each other” (*San Francisco*, n.d.). The following description emphasizes the moderation team’s specific interest in making r/SanFrancisco the absolute best it can be. Here, that means remembering that users behind the screen are also neighbors in a beloved city. Content that is hateful, violent, or otherwise treats individuals or groups as “second-class citizens or subhuman”. Meaningful engagement is defined as opposed to “just promoting”. Users are encouraged to “actually” join the community, which is further defined to be a “city of creators, artists, weirdos, and activists”. Despite the clear picture painted of who San Franciscans are, in the same breath moderators remind users to respect those with different points of view or perspectives. Content deemed worthy is content that has been created with effort. Effort as it is used by the moderation team of r/SanFrancisco as putting as much time into posting a question as one would expect

others to spend responding. In other words, the more context a post provided the more valuable it was considered in the eyes of moderators.

Ultimately, r/SanFrancisco was determined to be the most vague due to its reliance on buzzwords that implied meaning rather than providing clear instruction. While the guidelines were generally positive, they lacked specific explanations or enforceable rules. A closer look at the descriptions and language used reveals an underlying vision of the ideal San Francisco. Moderators also opted to hide the list of users who make up the team, lending further opacity to the moderation culture in r/SanFrancisco.

r/Austin

On the other end of the spectrum, r/Austin had more than double the number of rules with a total of 15. The first section in the “About” tab on r/Austin is a detailed “FAQ and Modding” section. This section consists of various supplementary links that lead to pages with additional information.

First, the moderation team links to a page of common questions that may be removed for repetition or lack of detail (*R/Austin FAQ*, n.d.). Included in this list are questions about moving, things to do in Austin, making friends, requesting further information about Austin-related incidents, lost and found pets, and asking why a post or user is getting downvoted. Also addressed are questions about posts regarding the Chili’s on 45th and Lamar, an inside joke within the subreddit that generally serves as a response to posts deemed low-effort. The moderators provide answers to each question, often including links to more appropriate places to find the desired information. The next bullet in the “FAQs and Modding section” explains to users that r/Austin is not the Craigslist—referring to the prohibition of posts seeking to buy or sell

goods or services. This bullet links to r/AustinClassifieds, redirecting those who might be looking to do so to the appropriate space.

The following bullet links to r/Austin Rules and Moderating Info, bringing users to a detailed webpage that outlines the fundamental understandings that guide the moderation team's actions (*R/Austin Rules and Moderation*, n.d.). This link addresses several r/Austin-specific concerns. First, the intentions of each of the mega-threads are defined. The weekly “Stuff to do in Austin” thread is for events or posts about events. Posts about pets that need homes or help go in the weekly “Pet Adoption / Pet Help Post” thread. Anything regarding real estate or moving go into the appropriately named thread, and “Weekly Stupid Question Sunday” is a thread where no question is too redundant or irrelevant. The moderators also define posts that have no business in r/Austin. They reiterate that posts looking to sell or buy belong elsewhere, including referral codes and items for sale or services for hire. Restaurant recommendations have their own dedicated subreddits, as well as personal requests or missed connections. Surveys, polls, petitions and market research are not allowed, and the moderators provide no alternative spaces within which to post them. Finally, moderators remind users that anything not specific to Austin, generally, belongs elsewhere— they recommend r/Texas.

This link also includes further reiteration of the r/Austin rules. Moderators specify that the rules in r/Austin are based on the platform guidelines set forth by Reddit— called “Reddiquette”. Both content and behavior are addressed. “Noise” comments, defined as being either single character or otherwise incoherent, will be removed. Personal information, either requesting it or providing it, will also be removed. This rule is also cited as the reason that cross-posts from other subreddits will also not be allowed. Other provisions include the deletion of posts with paywalls and limitations for crowdfunding efforts, which may be posted but only with

explicit permission from a moderator first. Trolling, or otherwise arguing in bad faith, disparagement, and wishing harm upon others are also warned against. The moderators provide an example for the latter of which: “He should be arrested for his crimes is alright” but “Somebody should put him out of his misery” is not. While most of these include citations to the Reddiquette page developed by Reddit employees, it is noted that hate speech— defined as racism, sexism, homophobia, or transphobia— is not explicitly addressed by Reddiquette but will “not be tolerated” in r/Austin. The penalty for doing so is not explicitly stated. Procedures are also outlined for how to deal with posts or requests regarding crime, emphasizing that any personal information included in the post is strictly prohibited. For posts regarding criminal cases to be approved the original poster must send in the case number or the name of the detective handling it for the moderators to verify it. It is requested that users familiarize themselves with all rules before posting.

The next section in the r/Austin Rules and Moderation link describes moderation responsibilities and practices in r/Austin. In r/Austin, these closely align with Reddit’s platform-wide suggestions for moderators listed within the Reddiquette. Moderators handle flair, special identifiable tags awarded to certain users for certain things. They can be created or removed by the moderation team. Moderators are also granted the responsibility of scheduling the weekly posts, removing posts that violate guidelines, and managing the general layout of r/Austin. At the end of this section, the following reminder is posted:

Please remember that /r/austin is not a place where we, the mods, try to decide who gets what voice, but rather keep it “all things Austin” - the key to deciding what is good and bad lies in that little up/down arrow to your left.

This explicit declaration of intention, referring to the ability of users to upvote and downvote content based on their perceptions of its alignment with community goals, further lends to the specificity present within r/Austin's moderation practices. Beyond this, sensitive topics are given their own section within the "r/Austin Rules and Moderating Info". Under this general subject, political posts are addressed first. Discussions of politics are allowed and encouraged, but the importance of relevance is key. They must be directly relevant and specific to Austin in that they have a direct impact on Austin, not the state as a whole. Those posts are for r/Texas. Articles cannot editorialize headlines either. If it is determined that the title of the post does not match the original headline close enough, the post will be deleted. Trolling, defined as repeatedly posting or deliberately provoking reactions, is permitted to an extent. Finally, and notably if only for its seemingly stark contrast to the general subject matter of the other guidelines, submissions that complain about specific dog breeds would not be allowed. While it is impossible to know exactly why this addition was made, this addition is specific enough to warrant notice for its apparent purpose in responding to an Austin-specific need. None of the other subreddits mention discussions of dog breeds, especially as a political subject.

Other topics deemed sensitive include posts about homelessness. Calls for violence or harm towards those who are homeless will be removed and the poster will be banned. Posts that the moderators do not feel contribute towards meaningful discussion, such as open-ended questions with no further context. The example "What can we do about homeless people" is provided. Finally, posts regarding missing persons must be verifiable. The moderation team provides no further explanation, instead linking to the rules of a subreddit called r/MissingPersons (*r/MissingPersons*, n.d.). Here, verification is determined through the inclusion of trustworthy sources in posts regarding the subject. According to r/MissingPersons, a

trustworthy source can be any news website, law enforcement website, missing person organization. Social media posts will be considered on occasion. These same guidelines are upheld in r/Austin, as per the “Missing persons” section of the r/Austin Rules and Moderating Info tab.

The posting of petitions is the last subject addressed in the “r/Austin Rules and Moderating Info” page. Petitions, including those that encourage users to sign up for or attend events supporting a cause, are prohibited in r/Austin. This marks the end of the topics addressed in the “r/Austin Rules and Moderating Info” page.

This page, included within the r/Austin “FAQs and Modding” section of the subreddit’s “About” tab, provides a detailed framework for content regulation, user behavior, and moderator responsibilities within the subreddit. It references sources from within r/Austin, from other subreddits, from official Reddit platform pages, and from sources beyond Reddit like other social media platforms and news sites. This page is centrally located within the subreddit, making it easily accessible to users.

The next bullet point under the “FAQs and Modding” section of the About tab in r/Austin addresses the use of AutoModerator in r/Austin. The moderation team uses AutoModerator to report racial slurs, add flair, schedule posts, and apply additional scrutiny to known trolls. However, the exact information and language used to train the AutoModerator configuration in r/Austin is not stated.

The rest of the “FAQs and Modding Section” reminds users that moderators always exercise best judgment and that exceptions will be made at their discretion. Any questions about this process are to be sent directly to the modmail— an interface that enables private messaging where users can send messages directly to the moderation team. While all users can send private

messages to any other users across Reddit, this addition provides a specialized interface and clear instructions as to how to proceed.

The next section in the “About” tab of r/Austin is the list of rules— 15 in total. This is the most of any subreddit selected for this study. The first addresses the use of personal information in posts. This is defined specifically as names, phone numbers, license plates, links to social networking sites and is strictly prohibited. Referencing another user's real name in the subreddit is also not allowed— even if they have previously disclosed this information. The exact penalty for this action is not explicitly stated.

The next rule is labelled “FAQ/Moving to Austin/Jobs in Austin”. While the subject matter addressed under this rule is wide, the sentiment is the same: “r/Austin” is not your personal Google.” Moderators heavily encourage users to do a separate search first, then come to Reddit with specific questions. In some cases, proof that the user has done this will be required for the post to remain. r/Austin also has a separate rule for “Real Estate/ Apartments Questions”. The general subject matter and tone are the same as that of previous provisions surrounding moving. It is once again reiterated that R/Austin is not the place to ask generic real estate questions like where to live or reviews of apartments. The moderators suggest that users looking for this information contact an apartment locator or a real estate agent for those looking to purchase. They also link to the most recent weekly housing thread and explain that posts will be removed and redacted there.

The next rule in the guidelines lists ads, spam, promos, and referral codes. Posts that fall under this section refer to promotions for any business, bar, event, Facebook group, or meetup. The posting of any promotions that may include compensation or incentives will be removed, including for bars or concerts.

The rule titled “Food/Gardening/Beer” questions refers to general entertainment and recreation posts with pre-existing, active subreddits that specialize in the area being referred to. For example, questions about food, restaurants, happy hours, and coffee shops belong in r/AustinFood. Similarly, gardening questions are for r/AustinGardening and r/AustinBeer handles all questions regarding beer. While the intention behind the creation of these subreddits and their inclusion within the guidelines is hard to tell, they are notably specific to r/Austin.

Harmful content was generally addressed under one rule— headlined “Habitual Negative Content/ Personal Attacks/ Racism/ Bigotry.” Users who post when subreddit karma is negative will be banned or otherwise removed by moderators as they see fit. Racism and bigotry are plainly not tolerated, offenders will be banned, and moderators explicitly state that new accounts or those not active on r/Austin will be treated with extra caution until proven trustworthy. Moderators do not explicitly state how trust is proven. The only other point asks users to use common sense, don’t personally attack people, and “basically, don’t be an asshole.”

Despite being addressed in multiple other areas of the “FAQs and Modding” and “Guidelines” sections within r/Austin, there is a rule labelled specifically “Jobs in Austin”. It is once again reiterated that r/Austin is not a job forum and that questions regarding any aspect of employment should be directed to r/AustinJobs. If not, these posts will be removed.

Within r/Austin’s extensive list of rules, the expectation that all content must remain directly relevant to the city is reinforced in multiple ways. The next rule in the list is titled “Keep it Austin Related” -- a phrase that is repeated often throughout and serves as an overarching directive for ensuring that posts maintain connection to the local community. The attitude here is much the same as previous sentiments expressed throughout the guidelines, with some additional requirements. Posts that fail to meet this requirement, whether they reference broader state-wide

issues or national matters, are subject to removal. While this rule is relatively straightforward, its enforcement is further refined in specific provisions regarding event-related discussions and political discourse.

Posts about concerts, meetups, and other local events are not permitted as standalone submissions and are instead redirected to a weekly pinned post. The same restriction applies to the city's two major annual festivals—SXSW and ACL—which, despite their international recognition and deep ties to Austin's identity, are deemed off-limits for general discussion within r/Austin. Instead, users are expected to direct their inquiries to r/SXSW or r/ACLFestival, both of which cater specifically to these large-scale events. However, an exception exists for news coverage; an article about an event may be permitted, so long as it is more than an advertisement and offers substantive information.

The guidelines also draw explicit boundaries regarding illegal activity. Any post involving the buying or selling of illegal drugs results in an immediate ban, without warning. This zero-tolerance approach is one of the most severe policies within r/Austin's moderation framework, reinforcing the subreddit's strict stance on illicit transactions. Similarly, commercial activity is strictly prohibited under the rule titled "This is not Craigslist", which prohibits posts seeking to buy, sell, trade, or hire for services. Users looking for such exchanges are instead redirected to r/AustinClassifieds, a separate subreddit dedicated to non-commercial transactions. The prohibition extends even further to requests for short-term borrowing or personal ads, reinforcing the subreddit's positioning as a space for discussions rather than commerce.

A similar philosophy is applied to reviews and recommendations under the provision "This is not Yelp." While asking for recommendations is permitted, posting reviews of businesses, restaurants, or stores is not. The moderators take particular issue with generic food

photography, discouraging posts that showcase common or unremarkable dishes. Images of queso, a drink at Chili's, or a Lone Star beer at a local bar are specifically mentioned as examples of content that will be removed. For food-related discussions, users are redirected to r/AustinFood, which serves as the appropriate space for restaurant discourse.

One of r/Austin's most strictly punishable rules concerns political discussions, particularly those that extend beyond the city's borders. While politics is not banned outright, the guidelines mandate that posts must be specifically about Austin. Discussions of national politics belong on broader subreddits like r/politics, while Texas state-level discussions are directed to r/Texas or r/TexasPolitics. Any post that does not explicitly focus on Austin's local political landscape is subject to removal. The subreddit also enforces a rule against editorializing news headlines. If the title of a submitted news story deviates significantly from the original published headline, the post is removed, and users are asked to resubmit it with the correct wording.

One of the most detailed provisions in r/Austin's moderation policy governs discussions about homelessness, an issue of relevance to Austin's civic discourse. The rules prohibit any calls for violence or harm toward unhoused individuals, stating that such content will not only be removed but will likely result in a permanent ban. Posts advocating for a specific position without fostering meaningful discussion are also prohibited. The moderators emphasize that r/Austin is not a platform for ideological advertising, meaning that posts must contribute to a broader conversation rather than serve as one-sided declarations. Open-ended posts that lack substance, such as simply asking, "What can we do about homeless people?" without elaboration, are flagged as low-quality content and may be removed at the moderators' discretion.

Taken together, these rules illustrate a highly structured and interventionist approach to moderation in r/Austin. Unlike subreddits that emphasize broad, loosely defined expectations, r/Austin's guidelines are explicit, detailed, and heavily rule-based. The moderation team prioritizes clarity and predictability, ensuring that users understand not only what is permitted but also why certain content is restricted. The subreddit's extensive use of redirection, whether to other subreddits, pinned threads, or designated discussion spaces, demonstrates a preference for organizational efficiency, reflecting a community governance style that is both structured and transparent.

r/Chicago

In r/Chicago, the about tab includes the simple description "Chicago, Illinois" (*r/Chicago*, n.d.). There are only three sections. The first is a list of rules for r/Chicago. There were nine rules total, the descriptions of which were highly detailed. Most of which included specific examples for further clarity. r/Chicago's rules provide a middle ground between specificity and flexibility—establishing firm prohibitions while also allowing for moderator discretion.

The first rule in the list, setting the tone for the moderation culture in r/Chicago, is the principle that all posts must be specific to the city. While this is a common feature of location-based subreddits, r/Chicago explicitly discourages content that is even only loosely associated with the city. A post does not qualify as relevant simply because it references a Chicago-based individual, organization, or sports team. It must impact the Chicago community.

The second rule in the list details r/Chicago's approach to dealing with harassment. Bullying, baiting, and public shaming may result in a ban. Digging through a user's post history to escalate an argument is explicitly warned against by the moderators, as is "being a dick in

general”. This detailed articulation of what qualifies as harassment suggests an active enforcement strategy. The provision of an example sets it out from other subreddits, which just refer broadly to intentionally provoking other users without providing any clear parameters for what might constitute such behavior. Bigotry and hate speech are also addressed under this rule. The moderation team makes it clear that bigotry and hate speech are met with immediate and severe consequences, most often resulting in bans.

r/Chicago also asks users to adhere to site-wide rule and refrain from any actions that would “break” the site. Users are explicitly warned against voter manipulation, subscriber fraud, brigading, and impersonation. Moderators also warn against doxxing and the use of the subreddit for communicating about distributing illegal drugs or fake IDs is specifically warned against. Notable here is the inclusion of two very different types of behavior. Doxxing, the act of posting the personal information of someone on the internet for the purpose of causing harm or otherwise harassing them, is addressed by other subreddits as harmful behavior— akin to trolling. r/Chicago is the only subreddit in which doxxing is mentioned in the same breath and liable to the same penalties as criminal activities like the distribution of illicit materials and fraud. The r/Chicago moderation team treats all these behaviors the same, warning that posting such content will result in a ban. In this rule, a link to the Reddit site-wide rules are also provided. These are different from Reddiquette, which is a more informal collection of values. Adherence to Reddiquette is suggested by the platform, but compliance with site-wide rules is mandatory.

All posts must also be tagged with the appropriate flair. NSFW content must be labeled as such. While the requirement to use flair is a common feature of many subreddits, r/Chicago positions it as a fundamental expectation of the users—not the moderation team.

Commercial activity and self-promotion are similarly restricted, reinforcing r/Chicago's identity as a discussion-based community rather than a marketplace. This is in line with the guidelines present in other groups as well. The prohibition of classified-style posts applies to buying, selling, crowdfunding, and job-seeking content– with explicit redirections to r/ChicagoList, r/ChicagoApartments, and r/ChicagoJobs. While lost items are generally prohibited, exceptions are made for missing persons and lost pets.

Another aspect of r/Chicago's moderation strategy is its expectation that self-posts must contribute to meaningful discussions. r/Chicago's approach here is consistent along all other subreddits. Moderators request that users put effort into their posts. The moderation team reminds users that anything easily answerable by a quick Google search or otherwise deemed low-quality will be removed or redirected to weekly threads or other appropriate subreddits. r/Chicago also discourages duplicates posts within a one-month timeframe. Users are expected to check existing discussions before posting to avoid redundancy and foster more diverse conversations. While multiple perspectives on the same news topics are permitted as sources from different outlets, identical links are not.

The approach to self-promotion within r/Chicago is more flexible than outright prohibitions found in some subreddits. Rather than banning promotional content entirely, r/Chicago permits users to share events or groups if they are active participants in the community. However, even engaged users are limited to one promotional post a month and are expected to participate in discussions beyond self-promotion. This approach is flexible on two accounts: it is easier on promotional content than other subreddits while still allowing for a great amount of discretion in determining who exactly is allowed to post such things. “Active

participants” is not quantified in any real way, it is up to the moderators to determine how active one must be to post promotional content.

Notable to r/Chicago’s moderation infrastructure is the provision it makes for crime. While crime is touched on in other subreddits, r/Chicago is the only subreddit that had a rule labelled specifically “Crime”. The moderation team discourages posts that do not have “wider impact” on the city. Certain types of crime-related posts are subject to removal. This includes sensationalized “crime recap” articles that summarize weekly violence, reports on individual acts of violent or petty crime that do not extend beyond private incidents, and discussions about statistical trends. Additionally, any posts that frame crime in a vague or inflammatory manner—particularly those that are aimed to “rile up users” rather than inform or discuss solutions—are likely to be removed.

Oriented among the subreddits that have examined thus far in this discussion, r/Chicago falls between r/Austin and r/SanFrancisco on this subject. While r/SanFrancisco does not address crime at all, r/Austin moderators not only define what they consider crime but require very specific information about the crime for it to be posted. Unlike r/Austin, which provides extensive supplementary documentation explaining why specific rules exist and how moderation operates/ r/Chicago does not offer an equivalent FAQ or moderation guide. While the rules themselves are clearly delineated, the enforcement process is left more implicit— meaning users are given boundaries but not necessarily insight into the logic behind every decision. This in contrast with r/SanFrancisco, where moderation decisions appear guided by a loosely defined cultural identity rather than explicit structural rules. By comparison, r/Chicago provides firm guidelines that define expectations in plain language.

Ultimately, the moderation practices of r/Chicago exhibit a level of transparency that is both highly structured but also carefully measured—positioning it between highly interventionist models like r/Austin and more open-ended, interpretive frameworks like r/SanFrancisco. While r/Chicago provides clear rules outlining what is and is not permitted, it does not offer the same degree of explicit explanation regarding how moderation decisions are made that some other subreddits do. Instead, much of its governance structure relies on a combination of strict content prohibitions and discretionary enforcement.

r/Philadelphia

Philadelphia had a list of 10 total rules. While the descriptions were brief, the language was very specific and direct. The moderation policies of r/Philadelphia are structured to balance specificity with flexibility, offering clear expectations while allowing for case-by-case discretion in enforcement. Unlike r/SanFrancisco and r/Chicago, where rules either rely on broad moralistic framing or strict procedural enforcement, r/Philadelphia's guidelines strike a middle ground, providing defined policies while also incorporating conditional allowances and moderator judgment in ambiguous situations. Although less detailed than r/Austin, which outlines exhaustive procedural requirements, r/Philadelphia provides a more structured approach to content regulation than either r/SanFrancisco or r/Chicago.

At the foundation of r/Philadelphia's moderation system is the expectation that all users follow Reddit's global "Reddiquette" guidelines. While adherence to sitewide rules is common among subreddits, r/Philadelphia's explicit emphasis on these principles signals a commitment to civility and constructive engagement. This is reinforced by a strict prohibition against personal attacks, both direct and indirect, with the enforcement of these rules left to moderator discretion. Notably, posts classified as "serious" or those discussing sensitive topics are subject to a higher

level of scrutiny, suggesting an awareness that discussions of a certain nature require more active oversight.

The requirement that all posts be relevant to Philadelphia reflects a territorial specificity similar to that found in r/Chicago and r/Austin, ensuring that the subreddit remains anchored in discussions of local significance. However, r/Philadelphia's guidelines provide a conditional allowance for certain non-Philadelphia topics, such as discussions of "shore posts," which refer to content about regional shore destinations commonly visited by Philadelphia residents. Statewide politics and government affairs are explicitly prohibited unless they have a direct impact on Philadelphia itself. This provision distinguishes r/Philadelphia from r/Austin, which permits some state-level discussions, and from r/SanFrancisco, where regional identity is more fluidly interpreted. Additionally, users who are uncertain about whether their content qualifies for inclusion are encouraged to contact moderators, demonstrating an engagement-oriented approach to content filtering rather than a strictly punitive one.

Like r/Chicago and r/Austin, r/Philadelphia enforces a strong distinction between discussion-based content and commercial activity. Self-promotional posts, solicitations, and crowdfunding efforts are not allowed as standalone submissions. Instead, local artists, musicians, and small businesses are given a designated space in the weekly "Small Business Thursday" thread, ensuring that such content is still represented within the community without overwhelming the main feed. Additionally, articles from legitimate news sources about local businesses or artists are permitted, distinguishing professionally published content from self-promotion. This policy is less restrictive than r/Austin's outright prohibition on many forms of promotion but more structured than r/SanFrancisco's flexible enforcement of commercial content.

The expectation that historical or archival images be sourced correctly further underscores r/Philadelphia's emphasis on verification and information integrity. Unlike many subreddits, where image-based posts circulate freely, r/Philadelphia requires either a direct link to the original source or an accompanying explanatory comment. This same standard applies to crime-related content and missing persons reports, which must be backed by a credible source such as a news article or police report. The explicit rejection of social media posts as valid sources—except in rare cases of active public safety emergencies—indicates an effort to mitigate the spread of misinformation. By contrast, r/SanFrancisco and r/Chicago do not impose such strict evidentiary requirements on crime-related content, allowing for broader community discussion without formal verification.

In keeping with common Reddit moderation practices, r/Philadelphia strictly prohibits advocacy of violence, hate speech, and discriminatory content. However, unlike r/Austin, which provides specific examples of what constitutes harm-inciting speech, r/Philadelphia's rule remains broad, leaving enforcement to moderator discretion. Similarly, r/Philadelphia enforces restrictions on editorialized news headlines, requiring users to post articles with their original titles rather than injecting personal opinion into the headline itself. This policy mirrors r/Chicago's prohibition on headline alterations, reinforcing a platform-wide expectation that news discussions remain as neutral as possible in their framing.

A notable feature of r/Philadelphia's guidelines is the restriction on "low-effort content," which includes duplicate submissions and overused memes. Unlike r/SanFrancisco, where engagement is often shaped by local in-jokes and informal cultural markers, r/Philadelphia explicitly discourages the repetitive circulation of popular memes or unoriginal content. A three-month restriction on posting identical or closely related content ensures that discussions remain

fresh and avoid redundancy. This standard is stricter than that of r/SanFrancisco, where meme culture plays a larger role in community identity, but aligns with r/Chicago's emphasis on preventing excessive repetition.

Moderation practices surrounding NSFW and graphic content in r/Philadelphia follow a similar logic of controlled permissibility. While sexually explicit material, gore, and violent imagery are broadly prohibited, news sources containing such content are conditionally approved. Unlike r/Austin and r/Chicago, where NSFW content is tightly restricted and often removed outright, r/Philadelphia allows for exceptions based on editorial judgment, permitting such material when it serves a broader journalistic or public interest purpose.

r/Philadelphia's moderation framework is highly structured but not rigid, offering a defined set of rules with room for discretionary enforcement in specific cases. While it does not reach the exhaustive level of detail found in r/Austin, it is significantly more explicit in its policies than r/SanFrancisco or r/Chicago. The requirement for credible sources in crime-related posts, the conditional acceptance of certain non-Philadelphia content, and the structured approach to self-promotion distinguish r/Philadelphia as a community that values accuracy, engagement, and localized discourse. The subreddit's guidelines prioritize information integrity, discourage redundant or low-effort content, and ensure that moderation decisions align with a balance of clarity and adaptability.

Summary of community guidelines

Ultimately, each subreddit had its own distinct approach to moderation, shaping community expectations and identity in unique ways. r/Austin had the most extensive and specific moderation framework and r/SanFrancisco the most open-ended. Meanwhile, r/Chicago and r/Philadelphia occupied a middle ground. Across all four, certain trends emerged.

Self-promotion was widely restricted. Posts promoting businesses, crowdfunding, or personal content were generally prohibited. It was common practice to redirect this content to other, dedicated subreddits— although the ease of which these were found varied across subreddits.

Community relevance was another foundational expectation. It was widely important that the discussions remained locally grounded, but once again what this constituted varied in definition across all subreddits. In general, content matter rules took the same tone but addressed different specifics. Some provided examples, while others did not. Specifically, crime-related content was handled differently across subreddits. While r/Chicago and r/Philadelphia imposed specific restrictions on discussions of crime statistics and individual incidents, r/Austin and r/SanFrancisco took a more flexible approach.

Finally, some moderators were much more accessible than others. r/Austin made the most provisions for structured communication channels through which to engage and resolve disputes. r/SanFrancisco and r/Philadelphia opted to hide the names of their moderators. These observed differences in moderation philosophy and enforcement mechanisms influenced the types of content shared and the ways users interacted within each subreddit.

Moderating moderation

While there were significantly less explicit discussions of moderation found during the selected time frame of this study than expected the incidents that were observed provide useful insight into perceptions of fairness and community identity. Users debated their interpretations of guidelines, their enforcement, and the implications on community interaction. Discussions of moderation took two general forms. Users either addressed the moderators directly by saying something like “Mods, deleted if not allowed”, then asking a question or otherwise addressing a

topic that might potentially violate guidelines. For example, a user in r/SanFrancisco posted to promote a subreddit they created for employees of the City and County of San Francisco. The user cited personally observing a consistent number of questions about working for the city or otherwise related to city employees as the reason for creating it. At the time of data collection, this post had not been deleted.

In Austin, a user with a Top 1% Commenter flair included the following disclaimer in their post: “Mods: I know this isn’t specifically Austin but so many of us are suffering right now that I hope it’s OK, and I used the allergy flair” (*Cedar Fever Hack?*, 2024)

The user then goes on to identify themselves as a “cedar fever sufferer”. They share a recently discovered home remedy for alleviating allergies, including a short discussion of existing scientific evidence to back their suggestion. As the user notes, the subject of allergies is not a challenge unique to Austin. However, oriented among the larger context of a significant number of posts within the time frame that also address cedar-fever allergies, it is apparent that this is an issue that significantly impacts the residents of Austin. In the comments, users shared their own remedies and commiserated on suffering with allergies—granting legitimacy to its proposed relevancy. It also fulfills requests to use the appropriate flair for posts.

More common, however, were more indirect addresses of the moderation team to encourage discussion of moderation practices. In r/Philadelphia, a user posted with the title “Serious Discussion Thread: The Mods are dorks here, care to elaborate anyone?” (*Serious Discussion Thread*, 2024). A comment from one of the moderators is pinned, in which they respond directly to “care to elaborate anyone?” with a succinct “No.” User responses indicate that this is generally understood to be a joke, though the moderator does not elaborate further.

The top comment is from a user who says: “Surprised they haven’t banned you for saying that.” Another says: “This sub has become very stale and boring. That is a fact.” Those who agree with this sentiment add that the space has become a news aggregator with occasional pictures, claiming it has become sterile and boring. Those who disagree argue that this is Reddit’s exact purpose, with one user even linking to Reddit’s Wikipedia page which describes it as a “social media aggregator site”. What is considered boring is also up for debate. Some users agree, claiming that r/Philadelphia used to be a much grittier place and should be again. Others disagreed. One user, responding to a comment complaining of boredom, says of the subreddit “Boring? It’s like being a visitor in an insane asylum where nobody is medicated.” When asked what r/Philadelphia “should” be, users cited a wide range of ideas. Others provided the alternative subreddit, r/philly, as an example of more relaxed moderation practices. One user described a “more casual atmosphere” that would be achieved by relaxing the rules to expand regulations on content. Other users provided more specific examples: the ability to talk about “tough topics” without posts being locked or deleted, a decrease or designated separate space for skyline and cityscape pictures. “Seeing a bird, the bridge and another skyline pic every god damn day is annoying AF. ‘Scroll past it’ doesn’t really address the fact that it brings absolutely nothing to a page that’s supposed to be topical issues and discussions,” for example. Others suggested the moderation team move away from relying on karma points to determine who is allowed to post because it alienates “real” users and visitors.

Multiple users cite the murder of George Floyd and the ensuing summer of riots and protests as the reason for implied change in moderation tactics. From their perspective, the space used to be much more tolerant of politically minded posts because there was a more even mix of serious and humorous, light-hearted posts. “Everyone wants to be a political pundit now,” one

user wrote. Others, however, argued that r/Philadelphia has always been like this and that it is like much of Reddit as a whole— even comparing r/Philadelphia to other location-based subreddits to provide context of how bad it can “really” get. Others argue the opposite, holding that issues experienced with moderation practices are unique to r/Philadelphia.

Ultimately, the sentiments expressed here are tolerant but skeptical of moderation practices. One user provides context:

I don't always agree with every decision they make, but when I've messaged the mods (politely), I've always gotten polite responses back that explain their thinking. That's enough for me. It's a volunteer gig and it seems like a really annoying one. Personally I'm not going to bitch about how someone does volunteer work that I'm not willing to do myself. You could volunteer to be a mod. You could start a new sub. You could spend less time on Reddit. So many options.”

Members of r/Philadelphia acknowledge the necessity of moderation while also addressing the difficulties of doing so well. Another user succinctly states “Fuck mods. Thank you mods.” Notably missing is any real engagement from the moderators themselves. Outside of two jokes posted by different moderators, receiving mildly positive but generally neutral feedback, moderators were not active members of this discussion in any real way.

In r/SanFrancisco, a user asks what might be done about the AutoModerator. In the body of the post, they explain that posts reported by users are now being taken down by the AutoModerator and not an actual member of the moderation team and express concern for what this might mean when users report posts for simply disagreeing with content (*What to Do about*

the Automoderator?, 2024). While this content would typically be restored by moderators, the user notes that they “appear MIA on that.” The user asks if anyone else has this problem and provides two solutions: either reconfiguring or removing the AutoModerator or expanding the moderation team. One user responds that the best option would be the latter, claiming “...automod is innocent, he’s not the issue. If they neuter him, our sub will just be flooded with spam.” Multiple users note that a call for applications to the moderation team was recently posted. However, those that applied claim they never heard back. Once again, the acknowledgment of the necessity of moderation is expressed alongside the displeasure of its current shape. One user, notably tagged with a “Top 1% Commenter” badge, says “The sub is full of ninnies. It is what it is”.

Other discussions of moderation in r/SanFrancisco garnered more engagement. One user posted to complain about the removal of a post for lack of relevance (*Mods Removed My Post about Nancy Pelosi Shafting AOC Saying It “Wasn’t Relevant,”* 2024). In the body of the post, the user explains that the original post was about how Nancy Pelosi “schemed” to prevent Alexandria Ocasio-Cortez from taking the Oversight Committee chair. The user claims this is as a strange intervention after what they perceive to be years of inaction in the face of “actual” conflict. The post received 219 comments. The ensuing discussion includes two general, broad trains of thought: users are either discussing whether Nancy Pelosi’s general position as a political representative of San Francisco is enough to make any news of her relevant to r/SanFrancisco or sharing their opinion on the political sentiments expressed by the original user. This would not have passed the guidelines requirements in more specific subreddits like r/Austin for its lack of direct impact on the community. However, the ambiguity inherent to r/SanFrancisco’s guidelines create gaps within which discussion arises. User opinions on

relevance generally leaned towards support for the original poster's comment. The top comment, supports the sentiment of the original poster that the content should not have been deleted:

"Nancy is SF's representative. Totally agree that the actions of our representative are relevant to her constituents." Another commenter supports, commenting the following: "Is this national news? Yup. Is it local news? Yes, in San Francisco and NYC. Sorry it didn't stay up so I could disagree with your post."

Most of the comments expressing support did so by explicitly addressing the user and the content in question. However, others expressed support more indirectly by attacking the moderators directly. Those that did claim that the moderators of r/SanFrancisco have political biases that impact the way they engage with the community. Some sentiments to this effect include "The mods like many people here now are fascist." Another referred to the moderation team as "centrist simps" that are only interested in maintaining a good image of San Francisco by allowing only redundant but mundane content to remain. Although centrism and fascism represent vastly different political ideologies, users invoked both terms as insults— suggesting displeasure at any perceived interference by moderators. Conversely, those that disagreed with the original posts' removal argued that the information being shared was federal news, not local news. These posts were all downvoted.

Another user posted out of concern for the removal of posts made by Scott Wiener, a California State Senator who identifies himself on Reddit as u/scott_wiener (*Why Are Scott Wiener's Posts Being Banned Here?*, 2024). The body of the post explains that Wiener has consistently contributed detailed, high-quality posts about relevant issues, though it is not clear from the posts alone whether they address California issues or San Francisco issues. According

to the original poster, the most recent banned post from Wiener detailed the California housing crisis and potential solutions for it.

The pinned comment is from a moderator. They explain that the thread was taken down by the AutoModerator, but that it has since been approved. They also include the following as a reminder: “And it might seem like we mods aren’t around, but we are. This one must have just slipped past us.” This comment was met with significant, involved engagement. Some users asked for or provided further clarity on how AutoModerators operate. It sparked discussions on how to better optimize the AutoModerator and whether to remove it entirely. Others offered support to the moderation team. One user, who identifies themselves as a past community manager, explains the difficulties inherent to the job:

So, as someone who was a community manager for many years: it is very hard, VERY HARD, to find people who want to do unpaid grunt work on the internet, with their only reward being people telling them they should to (sic) more and better work.

This was the most common sentiment echoed throughout this particular comment thread. In response to the approval of the post in question, the original poster attempts to engage with the moderator again by asking about the other 23 of u/scott_wiener’s posts that are currently banned or otherwise under review. Another commenter responds “we got our one instance of mod attention this year. seems like if we want any replies we’ll need to wait until sometime in late 2025 🤔”. This attracted the attention of another moderator, who reiterates that the moderation team are all volunteers and that they are always looking for new members. To apply, individuals simply need to attend a meetup, introduce themselves to the moderation team, and express their

willingness to contribute. Rather than a formalized process, the approach relies on in-person interaction, allowing prospective moderators to demonstrate their commitment through direct engagement with existing team members. This information is not posted anywhere in the guidelines.

The remainder of comments in this thread, mostly in response to comments posted by the moderators, express an extreme lack of dissatisfaction and general confusion regarding the moderation practices—highlighting a lack of consistency in claims by moderators that they are actively accepting new members of the team. Users claim that the moderators are notably inactive, such as the below.

...this is normal here bc we DO NOT have mods. instead, they've abandoned us to Automod, which is easily manipulated since it will pull down any post that gets too many reports. In fact, the last active mod who quit made a big farewell post about how the remaining mods REFUSE to add anyone else. Insanity,

Others question the level of censorship present in the subreddit, citing political motivations as the potential reason for the removal of a public figure's content.

Absolutely unacceptable from the mods. If we can't have open forum and discussion without censoring a public leader because we disagree with them, then what is this subreddit for? Just another online echo chamber? Get your shit together mods, stop being a part of what's wrong with this country.

Edit: Mods stated it was an auto-mod that removed the content. It's still concerning considering how popular the post was, and how his other post waited so long for approval. I hope they're able to find better tools and more people to keep this forum healthy, but I don't want to assign mal intent to something that was a mistake.”

While many users in the thread echoed similar sentiments, some questioned the authenticity of u/scott_wiener’s participation. One commenter argued, “He doesn’t discuss—he posts, uses this forum as a mouthpiece, and then never interacts here.” However, most responses pushed back against this claim, acknowledging that while he may not engage as frequently as others his posts consistently provide high-quality, relevant information. As a result, users generally agreed that his contributions, even if lacking direct engagement, remained valuable additions to the subreddit. Ultimately, the general sentiment expressed in this thread was negative towards moderation practices for its lack of rigor and specificity.

The high level of engagement on such posts suggests that this is an important topic to many r/SanFrancisco members. One week later, one of the moderators that engaged with the discussion of u/scott_wiener posted under the title “/r/SanFrancisco town hall: “Should public officials’ posts be exempt from flagging?”(*/R/SanFrancisco Town Hall*, 2024). The initial post about u/scott_wiener was linked, followed by a list of clarifications from the moderation team. The moderator explains that no one on the moderation team removed u/scott_wiener’s posts and that it was done because regular users were reporting the post. Moderators can restore a post if they notice excessive use of the report button, what they refer to as “report-button abuse”. In this case, however, it went unnoticed by the moderators. They describe the moderation inbox as a “firehose” and remind users that moderators are unpaid volunteers. Again, they extend an

invitation to apply. They then remind users that moderators don't make the rules, "you do"-- explaining that the standing rule in r/SanFrancisco is that one AMA (ask me anything posts, where experts offer their time and expertise to answer questions from users) per candidate per election was reasonable. The rule in question has been in place for 11 years, originally put in place after a vote that occurred when elected officials first started reaching out to r/SanFrancisco requesting to participate in AMAs. The moderator asks the following questions of subreddit members:

- When a public official makes a post here, should it be exempt from being taken down by the report button?
- Do we want to place any conditions on that privilege, such as requiring that they not just post submissions but also regularly jump into the comments? Or require them to first answer the horse/duck question?
- What should the maximum posting frequency be: once a day, once a week, once a month?
- Anything else I missed?

The sentiment expressed by users was unanimous. They felt that public official's posts should, by default, be exempt from any automatic action-- particularly those practices reliant on user action, such as flagging. Users expressed the desire to decide for themselves whether a public official was taking advantage of the space or otherwise disrespecting the community.

Totally agree. I want to hear from public officials, If they act like dicks they can be downvoted to oblivion just like any other redditor. They should be exempted from

automated removals, and shaming them is enough incentive for them to engage with the community. If they still don't engage, or use surrogates, that will speak for itself.

Users much preferred hands-on moderation in this case, citing the extreme nuance involved in making decisions surrounding content removal. One user added the perspective that, while public officials are not above the law, automatic moderation based on flagging was an issue for everyone. The solution was deemed to be the addition for more people on the moderation team. There was some inconsistency regarding the amount of people who have applied. Some users claimed they applied and never heard back, while others claimed that no one has or wants to. The decision to require an in-person meetup was discussed as a potential limitation, but claimed by some to be necessary in the interest of ensuring the person was local and “not some troll who lives across the country.”

Verifying a user's legitimacy was also an important part of the discussion. While users resoundingly agreed that posts from public officials should not be subject to automated moderation practices, there was debate over who constitutes a public official. Further, users discussed various means through which public official's accounts should be verified while still protecting the privacy of the users.

A week after the town hall, the same moderator announced that the moderation team had updated the configuration of r/SanFrancisco's AutoModerator to no longer remove public officials' posts due to user flagging ((*New Policy*, 2024)). They include public requests to public officials that they use the opportunity to engage with constituents, citing u/scott_wiener as an example. Additionally, the following stipulation is included: “Also, it will be official r/SanFrancisco policy that public officials are allowed, but not required, to tell us whether they

would prefer to fight one horse-sized duck or 100 duck-sized horses.” This is a reference to a comment made on the original post about u/scott_wiener that users referred to frequently throughout the discussion.

The moderator clarifies that a regular user’s post can still be taken down by the AutoModerator because of flagging prior to review by a moderator. To avoid this, they also include a link to apply to the moderation team. The link provided leads to a direct message forum with the subject “I Volunteer to Moderate”. While this post generated significantly less comments than the first two, those that did comment were generally supportive of the new rule. Many expressed their discontent with Scott Wiener as a politician but acknowledged his inherent contributions to the subreddit. Users cited the desire to hear all sides. This user states:

I'm a hater (forever pissed about the 2012 nudity ban he authored tbh), but I still want to hear what he has to say as our elected official! It's so *weird* to flag his posts like that, and I'm struggling to understand the rationale of my fellow haters.

Users were appreciative of the effort made by the moderation team to engage with the community and enact change based on the response. They cited this as an increase in transparency and a realignment of r/SanFrancisco’s community values with the expectations of its members. Overall, this discussion provides excellent insight into the processes through which users enact changes through collective, communicative action.

Content of discourse

Information requests

Users sought guidance, recommendations or general local insights. The nature of these requests varied widely, encompassing inquiries about local events, businesses, services, and real-time occurrences such as police activity, traffic disruptions, and noise disturbances. The high volume of such posts suggests that many users turn to these subreddits as their primary sources for community-specific knowledge and immediate answers to localized concerns.

While the topic of information requests varied widely in content, topics did not vary notably across different subreddits. It was not found that different subreddits had different information requests. While they were contextualized to the respective community, they were generally similar in content. Most consistently, these posts followed seasonal trends. The data was pulled at the end of the year, meaning many of the requests across all subreddits had to do with holidays like Thanksgiving, Christmas, Hanukkah, or Kwanzaa. Posters asked for location-specific gift suggestions and places to spend the holidays. For example, users in three of the four subreddits asked for suggestions on where to purchase a Christmas tree. While the fundamental intent to purchase a Christmas tree was the same across all posts, users added community-specific information to their requests.

In r/Austin, a user asked for a place to buy live Christmas trees and provided examples of similar local vendors they had purchased from in the past as added context for their request (*Live Christmas Trees in South/Central Austin*, 2024). In r/Philadelphia, a user asked for information regarding whether or not a local hardware store would be selling Christmas trees this year (*Fairmount Hardware Christmas Trees?*, 2024). A user in r/SanFrancisco asked for a Christmas tree vendor that balanced affordability with quality, with bonus points if they benefitted a charity (*Best Place to Buy a Christmas Tree?*, 2024). r/Chicago did not see any requests for Christmas

trees, which could suggest strict adherence to guidelines that prohibit buying and selling, inactive moderators, or simply a lack of need for Christmas trees in Chicago.

A significant portion of these requests focused on local events and activities, particularly recommendations for entertainment, festivals, and public gatherings. Additionally, there was a consistent demand for service-related recommendations. Users sought advice on contractors, legal professionals, medical practitioners, and specialized services tailored to their location. Public service inquiries, particularly those related to voting procedures and municipal governance, were also common— particularly given that the study’s data collection period coincided with an election cycle.

Inquiries about sudden or unusual local events represented another major category of information requests. These posts were most commonly phrased as “What was that?” or “Does anyone know what happened here?” and were relevant across all subreddits. Further information related to the city was often included in these posts, contextualizing them but also tying them into a larger observable pattern within subreddits and across them. For example, one user in r/SanFrancisco posted the following under the title “Ocean Avenue 24hr Fitness Incident” (*Ocean Avenue 24hr Fitness Incident*, 2024): “Anybody know what was going on around Ocean Avenue & Jules around 5pm today, 11/19/24? It looked like police and firemen were going into the garage of the 24hr fitness on the corner there but I'm not sure.” The post received two comments, both from users who saw the same thing. One provided details about the incident. Another user added further context, stating that they had also driven by the scene after it had cleared.

On December 17th, 2024, multiple users in r/Austin posted about a police helicopter circling over downtown Austin. Some included screenshots from flight tracking apps, others

cited visibly seeing the helicopter in the sky (*Anyone Know Why a Police Helicopter Would Just Do a Lap around the City in the Middle of the Night?*, 2024; *Police Helicopter Circling Bouldin*, 2024). One commenter claimed that they heard gunshots in the general area being discussed around 4:00 p.m. that day but had not seen anything about the news. They did not confirm if this was the reason for the helicopter or not. Other commenters offered jokes or otherwise apathetic responses like “Helicopters fly, it’s kinda their thing.” and “who gives af”, signifying annoyance at the repetition of post content.

The level of engagement on information request posts was largely transactional, as responses tended to be direct, factual, and goal oriented. Unlike more subjective or discussion-driven posts, replies to information requests were brief and to the point— often consisting of concise recommendations, confirmations of shared experiences, or links to relevant resources. However, engagement levels varied significantly depending on the specificity of the request. Posts that included detailed context, such as specific location preferences or criteria for recommendations, tended to attract higher engagement than vague, broad inquiries.

Across all subreddits, information requests were rarely flagged as violations or removed. The relatively low level of moderator intervention suggests that these posts were largely self-regulated by the community, with subreddit members maintaining norms of engagement without the need for significant external enforcement.

Public interest

Posts categorized under “public interest” often focused on sharing critical information with the broader community. These posts addressed topics related to local news, governance, and general civic concerns. These posts functioned as forums where users shared news articles, personal observations, and updates on significant community events. They were the most reliant

on external sources. Users linked to local news articles, government reports, or advocacy organizations to provide context for their discussions. All four subreddits mentioned providing sources, though the type of content this was required for varied. Regardless, the provision of sources by users was observed in each subreddit. Despite their relatively low engagement, public interest posts provide an interesting channel through which relevancy can be observed. While information requests arise out of need by the poster, public interest posts are rooted in the decision by one that a piece of information would be relevant to all. These posts were often political in nature but lacked any sway of opinion. However, there was also some variation between subreddits.

In r/Austin, information shared by users most centered around growth and development initiatives in Austin. Posts about the opening and closing of businesses, the passing of new zoning laws, and the construction of new apartment complexes and housing developments were highly prominent in r/Austin. While these kinds of posts were not unique to r/Austin, they were much more popular than in any other subreddit.

In subreddits with more stringent moderation practices, the subject of public interest posts varied less. In both r/Chicago and r/Philadelphia, information about local politics was significantly more frequent than any other type of share. Posts in r/Philadelphia frequently included union-related content or critiques of local government actions. In r/Chicago, many of the political posts were related to the city's current mayor, Brandon Johnson. While discussions did occur in the comments of these posts, they did not consistently attract conflict. In r/SanFrancisco, public interest was more civically minded. Posts were most commonly those that shared information about lost pets or belongings and missed personal connections. r/SanFrancisco did not mention the provision of sources at all.

This type of content reflects the impact of moderation on shaping discourse, as it was the least reliant upon and receptive of feedback from other users. In other words, user comments on the relevancy of another user's decision to share information were infrequent. As such, moderators and self-moderation are likely more responsible for their management.

Culture

Cultural discussions represented the smallest category of posts but generated the highest levels of engagement across all four subreddits. These discussions provided insight into how users collectively negotiated their local identity, reinforced community norms, and engaged in shared cultural experiences. Many of these posts initially seemed unrelated to local identity, but gained significance as users engaged with them— adding context, personal experiences, and shared community perspectives. A defining characteristic of culture discussions was their ability to transform seemingly trivial content into meaningful local discourse.

For example, a post in r/Austin featuring a photo of a uniquely decorated car gained substantial attention, with users discussing its history, its frequent sightings around town, and its significance as a local icon (*1,750 Lights, 1,575 Pieces of Tape, 20 Hours Later....This Is the Result! Merry Christmas Everyone!*, n.d.). Similarly, in r/Chicago, posts about one user's preference for scarves led to extensive conversations about shared experiences, long-standing traditions, and humorous takes on surviving Chicago's coldest months (*People Who Don't Wear Scarves - You're Really Missing Out.*, 2024).

Culture discussions also provided an opportunity for community members to reaffirm their perceptions of community identity by expressing pride and excitement for local events. In r/Chicago, multiple posts discussed a contest in Humboldt Park that judged lookalikes of Chicago native Jeremy Allen White's lead role in HBO's critically-acclaimed show, *The Bear*.

Carmie, a line-cook in a Chicago restaurant, is a widely recognized character both among locals and nationally. The contest was inspired by a similar event in New York City, which judged lookalikes of New York native Timoth   Chalamet in Washington Square Park. In the weeks that followed, dozens of replica contests popped up in cities across the country.

While some of those posts were more informational in nature, sharing details and instances of national news coverage about the event, the majority of them were personal accounts of the contest itself. The top post was a series of high-quality photos from the event. The top comment states simply: “What a great/ weird thing”. Similar comments expressing this sentiment were notably in the majority, with statements like “God, I love our city!” and “This is just great”. Comments that expressed disdain, such as “This show is so cringe for Chicagoans” received responses that defended Chicago’s identity. Adjectives that signalled a shared perception of authenticity and alignment with the working class, like Carmie, were used consistently in these defenses.

Conversely, users also connected over their shared disdain for local figures. In r/Austin, a post addressing local graffiti artist “Stench” garnered significant engagement. In a post titled “An open letter to Stench”, the original poster says: “Hey Stench, Your shit sucks. You’re not an artist, you’re just tagging. I can also run around town and write my name on a building with a sharpie. Knock it the fuck off” (*An Open Letter to Stench : R/Austin*, 2024). While this aggressive tone and intentional targeting of a named individual potentially violates several of r/Austin’s detailed guidelines surrounding respect, harassment, and personal privacy this post was allowed to remain active by moderators. Commenters discussed the legitimacy of Stench’s art, comparing him to other local Austin street artists and notable figures. Some appreciated his work, others made jokes at his expense. In addition to shared opinions about art, users also

discussed perceived changes to Austin's culture as reflected through shifts in quality and style of new graffiti art in the city. One commented "As a recent transplant, I am disappointed by the lack of quality in the graf work. Do better, scumbags!". The original poster responded, "Welcome, I Assure you, we used to have respectable scumbags back in my day!". One Top 1% Commenter viewed Austin's graffiti tags as a result of the abilities and priorities of their local government:

Graffiti tags are like a force of nature, if you get rid of stench you invite in the BUSCAR. Its not like they can be stopped unless APD really wants to stop em, and they do a shit enough job of catching people who actually hurt people. You may as well be complaining about the tides...

Ultimately, the post was locked by the moderator– allowing it to remain visible in the feed but preventing anyone new from commenting. The explanation was provided in a stickied comment, permanently attaching it to the top of the post, is as follows:

"We generally remove posts about these people, since they are mainly just looking for attention through vandalism, we don't want to be a venue for that. With that said, this has run its course, and some users are witch-hunting who it is. Locking it for those reasons."

In r/Philadelphia, there was a political angle to cultural discussions that was not present in other subreddits. Particularly surrounding the day of the 2024 election, multiple posts referenced Philadelphia's political influence and identity. An amateur image of downtown Philadelphia on November 4th, 2024 posted with the title "The most important city in America" garnered particularly high engagement from subreddit members (*The Most Important City in*

America, 2024) . Topics of conversation varied widely in scope, but political identity was largely expressed through disparaging comments towards Donald Trump. One meme in particular, a collage from the long running show *It's Always Sunny in Philadelphia*, was posted multiple times (Figure 3).

Figure 3: *Political meme shared in r/Philadelphia*



Multiple commenters referenced Philadelphia's political history with pride. One remarked, "Philadelphia saved America from the rule of King George. Now it needs to save American from the rule of King Donald". Similar statements include "Philly was where America started. Tomorrow, Philly will play a role in America remaining alive." and "Kind of fitting considering it all started here". Others, however, viewed the city's political identity as a burden. A comment that stated "I cannot wait to go back to being a filthy shithole no one wants to think about. I'm really tired of living in the political center of the universe" received significant engagement, supported by sentiments like "Knowing we will be bombarded with those ads again

in 3 years time has me considering moving to another state.” Both sentiments were engaged with equally.

Comments that attempted to campaign for a certain politician or otherwise impart political influence over other users in r/Philadelphia were notably lacking. Users who did take the opportunity to engage in more serious, information-driven political discourse focused on the importance of voting at all. Very little reference was made in support of one party or another within these discussions. Notably, r/Philadelphia is the only subreddit that does not have a provision in its guidelines that addresses political content.

The overwhelming majority of culture posts, however, were those that simply shared images from around the relevant city. Some were recent, posted by users to reflect a current moment in time. Others were older, posted as a throwback. They also ranged widely in quality, from blurry phone images to high-quality photography. Images were posted by self-identified city residents and visitors alike.

In each city, these images were generally of the same few notable city locations. In r/SanFrancisco, posts featured images of the Golden Gate Bridge and the Sutro Tower. In r/Austin, pictures were mostly of the city’s various natural features— such as Barton Springs and Zilker Park. Users in r/Chicago shared images of the Chicago skyline, Wrigley Field, and the shores of Lake Michigan. Still, other images shared did not feature any notable landmark. A significant amount of images shared simply depicted a sunset, sunrise, interesting cloud formation, or unique animal sighting. Despite sometimes being low-effort and oftentimes being redundant, these posts went largely unmoderated. Their abundance does not necessarily signal popularity, as image posts were consistently brought up in debates about the perceived decline in quality of the subreddit community.

Summary of findings

The findings of this study reveal how content moderation and community identity interact within location-based subreddits, shaping not only the types of content shared but also the ways discussions unfold. Each subreddit constructs its own regulatory framework, defining acceptable discourse and establishing rules that reflect the values of its members. While all four subreddits studied—r/SanFrancisco, r/Austin, r/Philadelphia, and r/Chicago—operate within Reddit’s broader platform regulations, they differ significantly in the specificity and enforcement of their guidelines.

Among the four, r/Austin maintains the most extensive and detailed moderation framework, with an exhaustive set of rules covering content relevance, behavioral expectations, and posting limitations. Moderators enforce these rules with a high degree of intervention, often redirecting discussions to designated threads or other subreddits. In contrast, r/SanFrancisco takes a more open-ended approach, relying on broadly defined principles that prioritize a sense of local identity over strict regulatory enforcement. While r/Philadelphia and r/Chicago both maintain structured moderation policies, they strike a balance between specificity and flexibility. r/Philadelphia’s guidelines focus heavily on source verification and structured political discourse, while r/Chicago enforces firm content restrictions but leaves room for discretionary enforcement by moderators. Across all four, self-promotion is widely restricted, with posts promoting businesses, crowdfunding, or personal projects generally prohibited or redirected to other subreddits. Community relevance is a key expectation, though the definition of what qualifies as relevant varies across subreddits. Crime-related content is handled differently, with r/Chicago and r/Philadelphia imposing restrictions on discussions of crime statistics and individual incidents, while r/Austin and r/SanFrancisco take a more flexible approach. Moderator

accessibility also varies. r/Austin provides structured communication channels for engaging with moderators, while r/SanFrancisco's team remains relatively opaque, with moderators often absent from discussions about their enforcement decisions.

Despite these differences in moderation style, explicit discussions of moderation by users were relatively rare, occurring in only a small fraction of posts. However, when moderation was discussed, it often reflected broader tensions within the community. Users debated their interpretations of subreddit guidelines, the fairness of enforcement, and the broader implications of moderation on subreddit culture. In r/SanFrancisco, user complaints about moderation often centered on a perceived lack of engagement from moderators and frustration with automated content removals. When a post from California State Senator Scott Wiener was removed by AutoModerator, users pushed back, leading to a town hall discussion and a subsequent change in subreddit policy exempting public officials from automated removals. In r/Philadelphia, criticism of moderation practices revolved around the subreddit's perceived shift away from allowing grittier discussions, with some users arguing that the space had become too sanitized. While discontent with moderation was expressed across multiple subreddits, users also acknowledged the difficulties of unpaid volunteer moderation, with some expressing appreciation for the effort moderators put into maintaining their respective communities.

Across all subreddits, information requests were among the most frequent types of posts. Users sought guidance on local services, events, and real-time disruptions, often using Reddit as a primary source for hyperlocal knowledge. While the subjects of these requests varied, seasonal trends were evident. Many users asked for recommendations related to holiday shopping, events, and decorations, often providing community-specific context in their inquiries. Posts about sudden or unusual local events were also common, with users frequently asking about police

activity, traffic disruptions, or loud noises. Engagement on these posts was largely transactional, with responses providing direct answers rather than sparking broader discussions. The level of detail included in the original post significantly influenced engagement, with more specific requests generally receiving more responses. Moderation of these posts was minimal, suggesting that information-sharing is widely accepted as a core function of these subreddits.

Public interest posts focused on sharing news and civic information relevant to the local community, often linking to external sources such as news articles or government reports. These posts tended to generate lower engagement compared to cultural discussions, with many users acknowledging their relevance but not actively participating in discussions. While r/Austin's public interest posts were primarily centered on growth and development initiatives, r/Chicago and r/Philadelphia saw more politically focused content, particularly regarding local government and union activity. In r/Philadelphia, source verification was emphasized more heavily than in other subreddits, with users expected to provide credible links for certain types of content. Despite their potential to shape civic discourse, public interest posts were not frequently debated among users, and moderation was minimal beyond initial enforcement of posting requirements.

Discussions about culture, while representing the smallest category of posts, consistently generated the highest levels of engagement. These discussions often transformed seemingly trivial content into meaningful reflections of local identity, with users adding personal context and shared experiences to posts. In r/Austin, a post featuring a uniquely decorated car led to an extended discussion about local art and community recognition. In r/Chicago, a discussion about scarves evolved into a conversation about winter traditions and the city's cultural resilience. Posts about local events, particularly those with a humorous or unexpected twist, were also widely popular. In r/Chicago, a lookalike contest for the lead character in *The Bear* sparked city-

wide enthusiasm. In r/Austin users debated the artistic legitimacy of a graffiti tagger named "Stench," leading to broader discussions about changes in the city's culture and public space. In r/Philadelphia, cultural discussions often took on a political angle, particularly surrounding the 2024 election. Users engaged in self-referential discussions about the city's political influence, with some expressing pride in Philadelphia's historical role in shaping the nation and others lamenting the burden of national political attention.

A significant portion of cultural discourse across all subreddits revolved around images of the city, often featuring notable landmarks or everyday urban scenes. These posts, while frequently criticized as redundant or low-effort, were rarely moderated. Despite their simplicity, they provided a space for users to engage in shared expressions of civic pride and nostalgia. However, in discussions about subreddit quality, users often cited the abundance of image posts as a sign of declining discourse, arguing that they detracted from more substantive discussions

This study examined the relationship between content moderation and community identity within location-based subreddits. Specifically, it sought to answer three research questions: (1) How do moderation practices in location-based subreddits influence the discourse that occurs within them? (2) How do moderation practices and guidelines shape community identity in location-based subreddits? (3) To what extent does community identity inform moderation practices in location-based subreddits?

The findings suggest that while moderation does not dictate discourse, it establishes a framework within which users negotiate the relevance of content and the boundaries of community norms. Rather than serving as a rigid or top-down enforcement mechanism, moderation operates as a guiding structure that subtly influences how conversations unfold. It shapes the conditions under which discourse occurs by setting expectations, defining what is

permissible, and reinforcing shared values over time. However, moderation is not imposed solely from above; it also emerges organically from within the community itself. Users actively participate in shaping the norms of their subreddit, reinforcing certain behaviors and discouraging others through patterns of engagement, self-regulation, and collective interpretations of relevance.

CHAPTER 5

DISCUSSION

As major platforms like Meta and X increasingly experiment with outsourcing moderation to users, questions arise about why decentralized governance appears to function effectively in some digital communities but leads to breakdowns in others. A central goal of content moderation research is examining whether governance structures actively shape user interactions by reinforcing community norms or merely impose predetermined rules with inconsistent enforcement. Understanding this distinction is essential for evaluating how Reddit's model of community-driven moderation fosters sustained engagement while other platforms struggle with trust and enforcement.

Unlike models where centralized moderators impose uniform rules, Reddit's governance system allows individual communities to define their own standards, creating an interplay between formalized subreddit guidelines and informal user expectations. This aligns with research on networked publics, which suggests that online communities develop their own regulatory mechanisms based on shared norms rather than external oversight (boyd, 2010). In this context, moderation functions less as a rigid control mechanism and more as an adaptable framework that users navigate and negotiate.

Moderating and shaping online discourse

An analysis of the sampled subreddits revealed that, despite differences in moderation styles, discussions across these communities remained remarkably consistent (RQ1). Topics such as crime, housing, public transit, local politics, and cultural events dominated subreddit discourse, regardless of whether moderation was highly structured and explicit or opaquer and more reactive. The consistency of these themes suggests that user interests and local concerns, rather than strict enforcement of rules, primarily drive discourse. Even when subreddit rules restricted certain topics—such as prohibitions on broad political discussions—users frequently adapted their framing to align with guidelines while still addressing their concerns. This aligns with research on information gatekeeping, which suggests that digital communities are shaped not only by explicit policies but also by the ways in which users collectively interpret and negotiate boundaries of relevance (Bruns, 2011; Shoemaker & Vos, 2009; Welbers & Opgenhaffen, 2018).

The way public interest posts were moderated also influenced how information was shared across different subreddits. Some subreddits, such as r/Austin and r/Philadelphia, imposed strict sourcing requirements, requiring posts about crime reports, missing persons, or public health concerns to be linked to verified news sources or government websites. Others, such as r/SanFrancisco, allowed more leniency, permitting anecdotal reports from users if they were framed as personal observations rather than factual claims. These differences in sourcing policies impacted not only the type of information shared but also how discussions evolved in response.

This finding aligns with prior research on credibility assessment in online communities (Metzger et al., 2003; Metzger & Flanagin, 2013), which suggests that sourcing requirements influence how users perceive and engage with information. When strict sourcing rules are in place, discussions tend to center around institutional credibility and factual verification,

reinforcing a reliance on authoritative sources. In contrast, when sourcing rules are more flexible, information-sharing is driven by collective knowledge and lived experience, allowing for a broader range of discourse but also increasing the potential for misinformation. This dynamic demonstrates that while moderation does not dictate what is discussed, it does influence how certain types of information gain legitimacy within subreddit communities.

The role of self-Regulation and pre-Moderation

The relatively low number of removed posts supports the argument that direct moderation does not significantly dictate the overall trajectory of discourse within location-based subreddits. Of the 8,000 posts initially requested for this dataset, only 7,603 were retrievable, indicating that approximately 397 posts were removed at some point. While it is impossible to determine the exact reasons for these deletions within the scope of this study, the relatively low removal rate suggests that subreddit moderators intervene selectively rather than aggressively censoring discussions. Instead of direct, large-scale enforcement, moderation appears to function more as a guiding mechanism, shaping the conditions under which discourse unfolds rather than strictly controlling its content.

Beyond post removals, pre-moderation tactics—such as karma requirements, AutoModerator filters, and keyword-based flagging systems—complicate the extent to which direct moderation influences subreddit discussions. Many subreddit rules prioritize proactive filtering mechanisms rather than reactive intervention by human moderators. In practice, this means that content removal often occurs before users even see a post, effectively preempting potential violations rather than responding to them after the fact. While this strategy is efficient in reducing moderation workload and preventing rule-breaking content from gaining traction, it

also creates an opaque system in which users may not fully understand why certain posts fail to appear. The absence of transparency in these preemptive measures limits users' ability to contest or even recognize moderation decisions, reinforcing a form of governance that operates largely in the background.

This filtering process disproportionately affects certain types of users, particularly those who have not yet accrued significant engagement history on the platform. For instance, r/Austin enforces strict karma-based posting requirements, ostensibly to maintain high-quality discussions and prevent spam. While these measures succeed in elevating more established users and ensuring that contributions meet community standards, they also create barriers to entry. Users with lower karma scores—who may still be knowledgeable and valuable members of the offline community—are systematically excluded from participating. This aligns with broader concerns in digital governance regarding the unintended consequences of algorithmic moderation systems, which often favor established voices and reinforce existing power structures (Gillespie, 2018; Matias, 2019a). By privileging long-time users and discouraging new voices, these systems subtly shape the composition of subreddit discourse, favoring continuity over change.

Timing may also play a role in the level of moderation observed. The dataset was collected during the winter holiday season, a period when moderator activity may have been lower than usual. If subreddit moderators were less active, this could explain why fewer posts were removed than might be expected at other times of the year. However, this study also found that a significant portion of content regulation was handled by automated moderation tools and user reports rather than direct intervention by human moderators. Even in periods of reduced moderator oversight, the broader infrastructure of subreddit governance—comprising algorithmic filtering, community-enforced norms, and user-driven reporting—remains active.

This suggests that moderation functions as a decentralized system of governance, in which various forms of content control operate simultaneously, often without the direct involvement of subreddit moderators (Gillespie, 2018; S. Roberts, 2017).

Moderation as a performative act

While direct content removals were relatively infrequent, moderation still played an important performative role in shaping discourse by reinforcing communal identity and signaling behavioral expectations. The language used in subreddit guidelines, pinned posts, and moderator interactions served as a subtle but powerful form of social conditioning— influencing user perceptions of appropriate discourse even in the absence of strict enforcement.

Subreddit rules do more than establish behavioral expectations; they construct and reinforce dominant linguistic norms that define the boundaries of community participation. The idea that language serves as a gatekeeping mechanism is central to third-wave sociolinguistics, which suggests that speakers are not passive but act as stylistic agents, tailoring language to establish identity and social hierarchy (Eckert, 2012, 2018). In digital spaces, particularly on Reddit, linguistic conventions mark belonging, influence discourse norms, and determine who holds authority within a community (Androutsopoulos, 2013; Tannen & Trester, 2013).

This is particularly evident in how subreddits define themselves through explicit linguistic framing. For example, r/SanFrancisco's rules define its members as "artists, activists, and weirdos," constructing an implicit identity that signals belonging through shared cultural markers. The use of casual, self-descriptive terminology creates an ingroup identity that fosters cohesion while subtly excluding those who do not align with the subreddit's tone or assumed values. In contrast, r/Austin defines its community by exclusion, stating what it is *not*: "a job

board, Craigslist, or Google.” This framing does not just set functional limitations but establishes a discursive hierarchy—positioning certain discussions as valid while implicitly discouraging contributions that do not conform to these unstated norms.

The way language is used in moderation practices mirrors broader patterns of digital linguistic dominance, where control over terminology, discourse structures, and community-specific jargon plays a central role in determining who is recognized as a legitimate participant and who is sidelined (Soliman et al., 2019). Those who understand the implicit cultural references, inside jokes, and unwritten expectations of a subreddit are more likely to be perceived as valuable members, while newcomers—who may lack familiarity with these linguistic markers—face greater difficulty integrating (Bucholtz & Hall, 2005; Danescu-Niculescu-Mizil et al., 2013).

This dynamic is not just theoretical but observable in how users police language in digital communities. Community members often engage in self-moderation by correcting newcomers, enforcing subreddit-specific terminologies, and responding negatively to those who fail to adopt the dominant linguistic style (Page, 2018). In r/Philadelphia, for example, discussions about public safety frequently distinguish between acceptable and unacceptable ways of framing crime narratives. Users who use overly sensationalized or politically charged language are often dismissed or downvoted, while those who discuss crime within the subreddit’s preferred framework—using neutral language, referencing official statistics, or localizing concerns—are more likely to be taken seriously. This demonstrates how moderation extends beyond explicit rule enforcement and into the realm of discursive power, where control over language shapes the legitimacy of contributions and, by extension, the broader flow of information.

By reinforcing these linguistic norms, subreddit moderation aligns with Barzilai-Nahon's (Barzilai-Nahon, 2008) concept of networked gatekeeping, in which access to discourse is regulated not only by explicit rules but by community-enforced linguistic standards. This gatekeeping is not necessarily negative—many communities rely on shared linguistic practices to foster belonging and maintain a coherent group identity—but it does raise important questions about who is excluded and why. If a user does not understand the implicit cultural references or subreddit-specific jargon, they may struggle to have their contributions acknowledged, regardless of their expertise or local knowledge. This illustrates how moderation operates on multiple levels: it does not simply remove content but shapes discourse by constructing the very linguistic conditions under which participation is deemed valid.

Moderation as discursive power

The findings of this study suggest that while direct moderation may be minimal, its influence on discourse is substantial. Through both explicit rule enforcement and implicit linguistic conditioning, moderators and community norms shape what conversations occur, how they are framed, and who is considered an authoritative voice. By reinforcing dominant linguistic norms, subreddit moderation functions not just as content control but as a mechanism for shaping discourse and defining community identity.

These insights align with broader research on digital governance and linguistic capital, which suggests that moderation structures in online communities serve not just to regulate behavior but to enforce discursive boundaries that shape knowledge production, social inclusion, and participation (Squirrell, 2019; Tufekci, 2017). This performative aspect of moderation underscores a key characteristic of Reddit's decentralized model: while explicit enforcement

remains limited, users are conditioned to regulate themselves through the very language they use, reinforcing Reddit's model of community-led governance as both a regulatory and cultural force.

Community identity, moderation practices, and guidelines

Findings suggest that moderation practices in location-based subreddits do not unilaterally impose identity through direct discourse control. Instead, subreddit identity is shaped through ongoing negotiations between users and moderators over what is considered relevant, shaping both formal moderation enforcement and self-moderation by users. This dynamic answers RQ2 (How do moderation practices and guidelines shape community identity in location-based subreddits?) by illustrating that community identity is not simply dictated from the top down but is co-constructed through interaction. While each subreddit prioritizes relevance and community impact differently, the underlying principle of iterative identity formation is consistent across all examined communities.

One of the most prominent ways that moderation influences subreddit identity is through structuring engagement behaviors. By implementing karma-based posting requirements, in-person verification, and other participation thresholds, subreddits create implicit hierarchies where users who frequently engage in discussions gain influence over community norms, while passive observers remain on the periphery. On Reddit, the ability to effectively engage in subreddit-specific discourse—including adherence to moderation norms and unwritten behavioral expectations—becomes a form of symbolic capital that determines one's standing in the community.

Moreover, the language used in moderation guidelines reflects a subreddit's identity, reinforcing its implicit cultural framework. For example, r/SanFrancisco's rules frame expectations through positive, declarative statements, defining its members as artists, activists,

and “weirdos”. In contrast, r/Austin defines its identity through exclusion, emphasizing that the subreddit “is not a job board, Craigslist, or Google.” These linguistic choices not only set behavioral expectations but also serve as cultural markers that signal belonging, shaping how users perceive their participation. This aligns with research on language and social identity, which emphasizes that community norms are reinforced through linguistic signaling and implicit expectations of shared knowledge (Benson, 2009).

Despite variations in subreddit-specific rules, internet-native terminology was present across all examined communities. Terms like “troll” frequently appeared in moderation guidelines, yet moderators did not explicitly define what constituted trolling behavior, assuming that users inherently understood the term’s meaning. This expectation of digital literacy positions internet culture as the dominant framework within these subreddits rather than local or regional linguistic norms (Androutsopoulos, 2013). However, given the fragmented nature of digital spaces, users who engage with different online communities may interpret these terms differently. A new user unfamiliar with Reddit’s broader culture might struggle to discern whether their behavior aligns with the subreddit’s expectations, illustrating how linguistic capital functions as a gatekeeping mechanism (Bourdieu, 1991).

The case of u/scott_wiener in r/SanFrancisco highlights how language in moderation guidelines actively shapes perceptions of legitimacy within a community. The debate over whether Wiener’s frequent posts constituted engagement or self-promotion hinged on the subreddit’s ambiguous rules. While some users viewed his participation as a constructive contribution to community discourse, others saw it as an exploitation of a public forum for personal or political gain. The ambiguity of terms such as “self-promotion,” “spam,” and “engagement” in subreddit guidelines played a crucial role in shaping user interpretations,

demonstrating that language in moderation policies is not merely descriptive but also performative—establishing implicit expectations about who is permitted to participate and under what conditions. This aligns with Phillips’ (Phillips, 2019) examination of how digital communities use coded language and implicit norms to maintain cultural boundaries.

A similar dynamic emerged in discussions about AutoModerator’s effectiveness. While users generally agreed that automated moderation played a necessary role in filtering spam, concerns arose regarding its ability to handle nuanced content decisions. The reliance on user flagging to determine content removal meant that posts were sometimes removed based on subjective interpretations of subreddit rules rather than clear-cut violations. This debate underscores a broader issue in digital governance: the challenge of maintaining consistency in rule enforcement when guidelines rely on subjective language or assumptions of shared community understanding (Gillespie, 2018).

Another notable linguistic choice across multiple subreddits was the use of the term "asshole" in moderation guidelines. While seemingly informal, its inclusion carried significant implications for community governance and identity formation. The term itself is inherently subjective—what one person perceives as "asshole behavior" may differ significantly from another's interpretation. This reliance on broad, colloquial terms assumes a shared cultural understanding among subreddit members, reinforcing the notion that the community operates under an unspoken consensus regarding acceptable behavior (Danescu-Niculescu-Mizil et al., 2013). However, such assumptions can obscure the complexities of individual interpretation and lead to inconsistencies in enforcement, ultimately shaping the form that community identity takes.

Despite their location-based themes, the absence of regional dialects or city-specific slang in subreddit guidelines suggests that these communities are more reflective of broader internet culture than of the cities they represent. This finding answers RQ3 (To what extent does community identity inform moderation practices in location-based subreddits?) by demonstrating that, although location-based subreddits are structured around physical places, participation and identity are shaped primarily by digital fluency rather than local knowledge. The expectation that users will understand Reddit-specific discourse, terms like "troll," "lurker," and "shitposting", suggests that Reddit's platform-wide norms take precedence over hyperlocal identity markers. This aligns with Bourdieu's (1991) concept of linguistic capital, wherein mastery of dominant discourse conventions grants legitimacy within a given social space. However, this also introduces barriers to participation for those who may not be deeply embedded in digital culture. A user new to Reddit may struggle to engage with a subreddit's unwritten social rules, even if they are deeply connected to the physical city it represents.

This dynamic reveals a fundamental paradox in the structure of location-based subreddits: while these communities are ostensibly organized around shared geographic spaces, participation and understandings of identity are shaped primarily by digital fluency. Moderation practices—through the language used in guidelines and the behaviors they implicitly encourage—function less as a means of directly enforcing local identity and more as a mechanism for shaping the form it takes.

By determining what is considered relevant, who is best positioned to participate, and how discourse is shaped, moderation practices subtly privilege certain forms of cultural belonging over others. The result is a digital community where participation depends less on shared physical space and local knowledge and more on the ability to navigate the linguistic and

behavioral expectations of an internet-native environment. Rather than acting as direct gatekeepers of community identity, moderators establish conditions under which identity is shaped by users through everyday engagement, self-moderation, and adherence to implicit social norms. This reinforces a model in which digital fluency becomes a key determinant of belonging, illustrating how online governance structures shape community participation in ways that extend beyond explicit rule enforcement.

Community identity as a guide for moderation practices

Moderation practices and community guidelines do not function as one-way directives. While moderation establishes the initial framework for acceptable discourse and behavior, it does not operate in isolation from the perceptions of community identity held by users. Instead, moderation exists in an iterative relationship with the community, where norms are continually shaped and challenged through user participation. This dynamic aligns with research on participatory governance in digital spaces (Katz et al., 2004; Lingel, 2017), demonstrating that online communities do not passively accept imposed rules but rather negotiate, reinterpret, and contest them based on shared identity and expectations. Moderators may set explicit rules—such as prohibitions on self-promotion or requirements for sourcing—but the way these rules are understood, enforced, and contested is ultimately dictated by the users who make up the community. When moderation practices fail to align with user expectations, they are met with pushback, renegotiation, or outright rejection. This reinforces the idea that moderation is a reflexive process that depends on continued dialogue between users and governing structures rather than a rigidly applied system (Gillespie, 2018).

This dynamic was particularly evident in how users responded to perceived overreach by moderators. In r/SanFrancisco, for example, the reference to moderators as “dorks” was not

simply a casual insult but a rhetorical signal of shared frustration. This terminology, though informal, functioned as a mechanism for constructing an in-group identity, reinforcing the idea that subreddit members shared a collective discontent with moderation practices. By using a term that is both dismissive and humorous, users signaled that they viewed moderation actions as overly rigid, out of touch, or socially awkward—yet the effectiveness of this sentiment depended on an implicit assumption that others in the community would recognize and agree with its meaning. This aligns with research on digital communities as spaces where linguistic markers function as tools of in-group identification (Tannen & Trester, 2013). Users construct an “us” versus “them” dynamic not just through explicit rule-breaking but through the ways they collectively frame authority, positioning themselves as the legitimate voice of the subreddit in contrast to an externalized moderator presence.

One of the clearest ways in which community identity informs moderation is through the negotiation of relevance. Posts that initially appear off-topic or unrelated often gain legitimacy through community engagement, illustrating how norms evolve dynamically based on user participation rather than rigid enforcement. Consider the case of the car covered in Christmas lights in r/Austin. Given the subreddit’s strict guidelines regarding relevance, specifically the requirement that posts have a direct impact on the local community, one might assume that this post could be flagged as off-topic or redirected elsewhere. However, the comment section transformed the post from a seemingly random image into a significant cultural moment for the subreddit. Users reminisced about seeing the car in previous years, discussing its place in local holiday traditions and likening it to uniquely Austin phenomena. This engagement effectively legitimized the post, reinforcing Lingel’s (2021) assertion that digital community norms are

performative and shaped through collective meaning-making rather than simply dictated by formal rules.

A similar process unfolded in r/Philadelphia, where discussions surrounding the city's role in the 2024 presidential election provided a case study in how community identity informs moderation through negotiated relevance. According to the subreddit's guidelines, federal news was generally prohibited to avoid divisive political discussions better suited for broader political subreddits. However, in the weeks leading up to and following the election, posts about Philadelphia's influence in national politics proliferated. This apparent contradiction highlights how moderation is not an absolute determinant of discourse but rather a flexible structure that bends in response to collective user sentiment. Philadelphia's role as a Democratic stronghold within a historically contested swing state became a focal point for subreddit discourse, with users framing their participation not as detached spectators of national politics but as active contributors whose votes and civic engagement shaped electoral outcomes. This aligns with research on networked publics, which suggests that online spaces function as hybrid information infrastructures where traditional political boundaries are reconfigured based on localized digital engagement (Papacharissi, 2010).

Unlike posts simply discussing national candidates or policy, these election-related discussions were deeply tied to Philadelphia's local identity. Users shared experiences about working at polling stations, navigating increased media attention, and witnessing the impact of national scrutiny on their city. The discourse was less about abstract political ideologies and more about Philadelphia as a place—a city with its own political culture, historical significance, and civic responsibilities. The fact that moderators largely allowed these discussions to persist demonstrates the adaptability of moderation in response to collective determinations of

relevance, reinforcing the idea that governance in digital communities is not static but negotiated (Van Dijck, 2009). When a topic becomes deeply intertwined with community identity, users often push against formal rules to maintain the conversation, further emphasizing how user engagement—not just moderation—determines the lived experience of subreddit governance.

Further evidence of this negotiation process can be observed in posts that include disclaimers such as “Mods, delete if not allowed.” This phrase, commonly seen across subreddits, is an explicit acknowledgment of moderation while simultaneously testing the boundaries of what is permissible. The use of this disclaimer indicates that users are aware of subreddit guidelines yet believe there may be room for interpretation or flexibility. In r/Austin, for instance, a user posted a remedy for cedar fever, a seasonal allergy caused by mountain cedar pollen in Central Texas, accompanied by a “Mods, delete if not allowed” disclaimer. While a post about seasonal allergies might be considered off-topic in a general subreddit, within r/Austin, it resonated as a highly relevant local concern, illustrating how community identity actively determines the application of rules. Users engaged with the post by sharing additional remedies, discussing the severity of symptoms in different parts of the city, and reinforcing its cultural significance within Austin.

This negotiation aligns with research on participatory culture and digital affordances which emphasize that users do not merely consume content but actively shape the communicative norms of their platforms (Massanari, 2015). The cedar fever post highlights two key aspects of moderation in location-based subreddits. First, it demonstrates that users internalize guidelines and engage in pre-moderation by assessing the appropriateness of their content before posting. Second, it shows that subreddit relevance is not strictly dictated by predefined rules but by the evolving consensus of engaged participants. This supports research

on self-moderation in digital communities, where collective validation determines what is permissible even in cases where content does not strictly adhere to written policies (Centivany, 2016; Hampton, 2004).

These examples underscore the participatory nature of subreddit governance. Moderation does not exist in a vacuum. It responds to, and is shaped by, community identity. This reinforces the study's research question: To what extent does community identity inform moderation practices in location-based subreddits? (RQ3). The findings suggest that while moderators provide a structural framework, the actual enforcement and reinterpretation of rules depend on the lived experiences and interactions of subreddit users. When subreddit identity conflicts with rigid moderation, community engagement actively reshapes the application of rules, often leading to adjustments in enforcement practices. This iterative process exemplifies the co-construction of digital governance, where power is not solely in the hands of moderators but is distributed across a network of active participants (Benkler, 2007; Matias, 2019b).

Ultimately, moderation in location-based subreddits operates less as a static mechanism of control and more as a dynamic system of negotiation. Community members do not simply adhere to pre-established guidelines; they test, contest, and redefine them based on their collective understanding of identity and relevance. In this sense, moderation serves as both a regulatory structure and a discursive site where users continuously shape the boundaries of acceptable discourse through participation, interaction, and resistance. As digital spaces continue to evolve, these findings emphasize the importance of user-driven governance in sustaining meaningful, community-oriented online engagement.

Community identity as a check on moderation practices

This study suggests that users hold substantial power in shaping moderation through collective behavior, discourse, and resistance. While moderators enforce rules and set structural boundaries, the community itself ultimately determines the legitimacy and acceptance of those rules. Users push back against moderation when they perceive it as misaligned with the shared values and cultural identity of the subreddit, effectively using community identity as a check on moderation practices. This aligns with prior research on participatory governance in online communities, which emphasizes that digital governance is most effective when it reflects the evolving norms and expectations of its user base (Leibmann et al., 2025; Matias, 2019b; Reddy & Chandrasekharan, 2023). Instead of moderation serving solely as a top-down mechanism of control, this study finds that it operates through a process of continuous negotiation between users and moderators, reinforcing the idea that digital governance is a collaborative effort rather than a unilateral imposition of rules.

One of the clearest examples of this negotiation emerged in the way users expressed frustration with moderators by challenging their legitimacy. In the discussion surrounding Nancy Pelosi's relevance to r/SanFrancisco, a user criticized the post's removal by asserting, "the mods that removed the post probably don't even live in SF." This statement reveals an important dimension of how users evaluate moderation decisions—not just in terms of rule adherence but in relation to local authenticity. By challenging the moderators' local credibility, users implied that effective governance of the subreddit requires an intimate understanding of the city's political and cultural dynamics. This reflects research on place-based digital communities, where the ability to participate meaningfully is tied not only to platform norms but also to deeply embedded local knowledge (Lingel, 2021). In this framing, moderation is not simply a technical function—it is a form of cultural stewardship. The expectation that moderators should be true

locals to make informed decisions reinforces the role of place-based authenticity as a defining factor in community governance.

This phenomenon extends beyond political discussions. Users frequently express resistance to moderation decisions when they feel those decisions contradict the organic development of subreddit culture. For instance, when moderators remove posts that garner high engagement and align with perceived community values, users often question the validity of these decisions. This was evident in r/Austin's Christmas lights car post, where users collectively determined its relevance despite its seeming misalignment with the subreddit's strict rules. Similarly, in r/Philadelphia, posts referencing the city's role in the 2024 presidential election remained popular despite subreddit guidelines generally prohibiting federal political news. These instances highlight how user consensus, rather than strict adherence to guidelines, often determines what is considered relevant to the community.

The pushback against moderation practices that do not align with community expectations further illustrates the reciprocal relationship between community identity and governance. Moderators attempt to shape subreddit discourse through rule enforcement, but their authority is not absolute. Instead, moderation is a negotiation, where users actively redefine and challenge boundaries through engagement, resistance, and discourse. This supports prior research on networked gatekeeping and user-led governance, which suggests that digital communities function best when users feel a sense of ownership over the space they inhabit (Barzilai-Nahon, 2008; Fiesler et al., 2018b). As a result, the effectiveness of moderation is not solely determined by its clarity or consistency but by how well it aligns with the community's collective self-perception.

This study finds that community identity is not just a passive feature of location-based subreddits—it is an active force that informs, shapes, and constrains moderation practices. Rather than functioning as a static set of rules imposed by moderators, moderation emerges as a constantly evolving process that is deeply intertwined with how users collectively define their community. This directly answers RQ3, which asks how community identity influences the enforcement and reception of moderation practices. The evidence suggests that moderation is most effective when it is perceived as legitimate by the user base and that legitimacy is largely determined by whether moderation aligns with shared cultural expectations.

One of the ways that community identity influences moderation is through user pushback when moderation is perceived as misaligned with local culture. When users feel that moderation does not reflect their shared values or priorities, they express frustration and challenge the legitimacy of the moderators. This was evident in r/SanFrancisco, where critiques of moderation often invoked place-based authenticity to discredit decisions that users found problematic. The belief that moderators should be true locals—rather than outsiders—became a common refrain in discussions of removed content. In this context, moderation decisions were evaluated not only based on the formal rules but also on whether moderators themselves were perceived as authentic members of the community. By using local identity as a form of critique, users establish a check on moderation, reinforcing the idea that governance within digital communities must be representative of the collective identity it seeks to regulate.

Another way that community identity informs moderation is through the way users collectively determine what is relevant content, even when it contradicts formal rules. Although subreddit guidelines often set boundaries around acceptable topics, users negotiate those boundaries in real-time through patterns of engagement and discourse. This was particularly

evident in r/Philadelphia, where discussions about the city's role in the 2024 presidential election frequently emerged despite formal prohibitions on federal political news. While these posts technically violated subreddit rules, they persisted because the community collectively deemed them relevant. Users framed these discussions not as general political news but as reflections of Philadelphia's political identity within a swing state, reinforcing their local significance. Similarly, in r/Austin, a post featuring a car decorated in Christmas lights gained traction despite not explicitly aligning with the subreddit's strict content guidelines. Initially, this post might have been seen as off-topic, better suited for a subreddit dedicated to cars or holiday decorations. However, users provided context in the comments, reminiscing about seeing the car in previous years and emphasizing its role in Austin's holiday culture. By engaging with the post and recontextualizing its meaning, users effectively expanded the boundaries of relevance, demonstrating how community identity is not just reflected in moderation practices but actively shapes them.

Additionally, moderation practices evolve in response to community feedback, further illustrating the reciprocal nature of governance in location-based subreddits. While moderators establish initial rules and enforcement mechanisms, these guidelines are not static. They shift in response to ongoing discourse and user resistance. The controversy surrounding the AutoModerator system in r/SanFrancisco exemplifies this process. Users criticized the automated flagging system for suppressing certain types of posts, particularly those involving political figures. While the system was designed to enforce subreddit guidelines, users felt it was being applied in a way that stifled legitimate discussion. In response to significant community backlash, moderators adjusted the system, demonstrating that moderation is not simply imposed from above but is instead a negotiated process influenced by collective user sentiment. This

aligns with prior research on algorithmic governance, which suggests that automated moderation systems require continuous adaptation to align with user expectations (Gorwa et al., 2020). These findings contribute to the broader understanding of digital governance, reinforcing that moderation is not solely dictated by platform policies or subreddit moderators—it is a continual process of negotiation, contestation, and adaptation. Rather than a one-directional process in which moderators impose rules on users, moderation is a two-way exchange in which users continuously shape the standards that define their digital community. The process of determining what is acceptable is never fully settled, as community norms and identities are fluid and subject to reinterpretation over time. This directly addresses RQ2 by demonstrating that moderation is not simply an enforcement mechanism, but a social process shaped by user participation, engagement, and resistance.

Ultimately, the study finds that moderation in location-based subreddits is not merely about maintaining order but about maintaining legitimacy within a community-defined space. As digital communities continue to grow, understanding the interplay between governance and identity will remain critical in ensuring that moderation practices remain transparent, participatory, and reflective of the values that shape online spaces. Rather than viewing moderation as a top-down structure, this study suggests that it is an ongoing process of negotiation, where community identity is both shaped by and serves as a check on governance.

CHAPTER 6

CONCLUSION

This study examined the complex relationship between moderation practices and community identity within location-based subreddits, revealing an ongoing, dynamic negotiation between users and moderators. Rather than existing as a one-way system in which moderators dictate discourse, findings suggest that community identity and moderation practices inform and shape one another in an iterative process. Users push back against moderation when it does not align with shared community expectations, while moderators adjust rules in response to evolving user behaviors and norms. This reciprocal exchange highlights the fluid nature of digital governance and challenges traditional understandings of moderation as a purely top-down mechanism.

Big tech companies such as Meta and X have moved toward moderation systems that rely on user participation, with initiatives like Community Notes aiming to crowdsource content evaluation rather than enforce policies through centralized oversight. The logic behind these shifts reflects a broader trend in digital governance, where platforms seek to offload the burden of content moderation while simultaneously distancing themselves from accountability for decision-making. However, as seen in this study, Reddit's community-driven moderation does not function as an entirely decentralized or independent process. Instead, it relies on ongoing negotiations where users actively contest, reinforce, or reshape governance through lived engagement.

This study found no single determinant of subreddit governance—community identity informed moderation practices, and moderation practices, in turn, shaped community identity. This bidirectional influence creates a structure that is not rigid but adaptive, allowing subreddit communities to evolve over time while maintaining a sense of cohesion. As one user insightfully framed it:

Reddit tends to attract various kinds of personalities. I like to think of them as 'mods' and 'rockers' (to borrow a concept from Quadrophenia). The mods like order and like to smugly arrange things to their liking, and most importantly, think they're absolutely 'right' about their opinions. The rockers like to push the mods' buttons, and stretch the limits of what's acceptable, and fight against boredom.

This analogy encapsulates the ongoing push-and-pull between structure and resistance that defines subreddit governance. Rather than being dictated solely by moderators or an overarching platform, subreddit identity emerges from an ongoing dialectic, where norms are continually renegotiated. Moderators enforce rules, but users determine their legitimacy through collective agreement, contestation, or reinterpretation. Moderation is not simply an act, then. It is a conversation, a ritual. Unlike top-down moderation models that rely on automated decision-making or opaque content removal policies, subreddit governance is marked by transparency and accountability—features that other platforms have struggled to replicate.

This study also reinforces the importance of flexibility in digital governance. Subreddits that have persisted for over 15 years have done so not because of strict rule enforcement, but because of their ability to accommodate change. Shifts in platform policies, evolving user

demographics, and external social and political forces have influenced these spaces, yet their fundamental structures remain intact. This is because moderation within subreddits is not just a set of policies but a living, adaptable process that evolves alongside the community itself.

While much of the discourse surrounding digital governance focuses on either platform-wide policies or algorithmic intervention, this research highlights the need for further study into mid-level governance structures—such as subreddit moderation—that mediate between top-down platform policies and grassroots user participation. Understanding these models is critical as major platforms continue to experiment with user-driven governance mechanisms while maintaining corporate control.

By examining how location-based subreddits navigate moderation and identity, this study contributes to a larger conversation about the future of digital governance. As platforms continue to shift responsibility onto users, Reddit’s model demonstrates both the strengths and the limitations of community-driven moderation. While it fosters organic discourse and adaptability, it also depends on sustained participation, collective oversight, and a balance between structure and resistance. Moving forward, future research should explore whether corporate platforms can meaningfully implement participatory moderation structures without undermining their transparency and accountability, or if these models will remain unique to community-driven spaces like Reddit.

Limitations

While this study provides valuable insights into the interplay between content moderation and community identity in location-based subreddits, several methodological limitations must be acknowledged. These constraints, both structural and ethical, shape the interpretation of findings and highlight potential avenues for future research.

Restrictions on Reddit's API

One of the most significant methodological challenges in this study stemmed from limitations imposed by Reddit's Application Programming Interface (API). While Reddit presents itself as an open-source platform, the extent to which researchers can access data is tightly regulated. Reddit's API only allows for a maximum of 1,000 posts to be retrieved per request, meaning that the dataset used in this study was limited to the most recent posts from each subreddit at the time of data collection. Importantly, posts cannot be selected according to time frame, keyword, or other criteria; instead, only the most recent content is retrieved, regardless of its relevance to the study's research questions (Reddit, 2025).

Beyond post limits, rate restrictions further constrained data collection. Reddit imposes strict rate limits to prevent excessive traffic from overwhelming its servers, restricting the number of requests that can be made within a given time. While this is a reasonable safeguard against misuse, it also meant that data retrieval was not instantaneous and required staggered requests over time, potentially missing key discussions that occurred between retrieval intervals. Moreover, these limits introduced a form of selection bias, as the data disproportionately reflected the activity of more active and engaged subreddits while quieter communities with fewer posts were less represented.

Additionally, recent policy shifts on third-party data access further restricted researchers' ability to examine historical trends in subreddit discourse. Reddit has implemented tighter controls on external tools like Pushshift, a widely used service that previously enabled access to archived Reddit posts and comment histories. Pushshift was an invaluable resource for longitudinal studies of Reddit communities, allowing researchers to track how subreddit discussions evolved over time. However, in response to concerns that AI companies were using

these archives to train models without user consent, Reddit restricted access to Pushshift data, significantly limiting researchers' ability to conduct historical analyses of subreddit behavior and moderation trends (Naqvi, 2024). This policy change meant that this study could only analyze posts from the live API pull, preventing deeper historical comparisons.

While these limitations did not compromise the core research questions of this study, they underscore the evolving challenges of conducting digital ethnography in increasingly privatized online spaces. Future research would benefit from alternative data sources, longitudinal tracking methods, or direct partnerships with subreddit moderators to access a more comprehensive dataset.

Ethical considerations of data collection

While Reddit positions itself as an open-source platform where public data is accessible for research, the ethics of data collection in digital spaces remain complex. This study adhered strictly to Reddit's API usage guidelines, ensuring that all collected data originated from publicly available posts within location-based subreddits and was used solely for academic purposes. However, ethical challenges in digital ethnography extend beyond compliance with platform policies and require critical engagement with broader questions about consent, privacy, and data governance.

One of the primary ethical debates in digital research revolves around the assumption that publicly available content can be freely used for academic study. Zimmer (Zimmer, 2017) critiques this notion, arguing that "public" does not necessarily mean "intended for public analysis"—a distinction that becomes especially relevant when studying social media platforms where users may not expect their contributions to be systematically collected and analyzed. This aligns with broader discussions in internet research ethics, which emphasize that user

expectations, not just legal frameworks, should shape ethical decision-making (Markham & Buchanan, 2012).

The Association of Internet Researchers (AoIR) has repeatedly underscored the importance of contextual integrity when conducting digital research, emphasizing that researchers should consider how the original audience for online content might differ from its use in an academic setting (franzke et al., 2019). Although Reddit users post within public forums, their engagement is often situated within specific community norms—raising ethical questions about whether repurposing this data for external analysis alters its meaning or violates implicit expectations of platform-specific discourse.

boyd and Crawford (2012) further highlight the power dynamics of data collection, arguing that researchers must critically examine who benefits from the extraction and analysis of online content. While this study was conducted for scholarly purposes, these concerns are especially relevant given the growing use of user-generated data by commercial entities, including AI companies leveraging social media archives for model training (Hutton & Henderson, 2022). The increasing privatization of public discourse—whereby platforms exert greater control over data access while simultaneously profiting from user engagement—adds another layer of complexity to the ethics of studying digital communities.

Another ethical challenge is the long-term accessibility of Reddit content, and the potential risks associated with archiving user-generated data. Unlike ephemeral social media posts, Reddit threads remain publicly visible for extended periods and can resurface years after initial publication. This raises concerns about whether researchers should take additional steps to anonymize data, even when users have willingly posted under pseudonyms.

Vitak, Shilton, and Ashktorab (2017) caution that data persistence in digital environments complicates traditional research ethics models, as posts that seemed inconsequential at the time of publication may take on new significance in different contexts. While this study did not collect personal identifiers or usernames, future research may need to consider enhanced anonymization practices, such as paraphrasing user-generated content rather than quoting it verbatim to prevent re-identification through search engines.

Additionally, the shift toward platform-controlled data access has impacted researchers' ability to study online interactions over time. Reddit's recent restrictions on tools like Pushshift, which previously enabled researchers to access archived discussions, demonstrate the growing tension between platform governance, user privacy, and research accessibility (Naqvi, 2024). While these changes are largely aimed at preventing the misuse of social media data for AI training, they also pose challenges for longitudinal studies of content moderation and community evolution.

The evolving landscape of digital research ethics suggests the need for greater transparency and community engagement in data collection practices. While studies like this one operate within platform guidelines, best practices in ethical research increasingly call for participatory approaches that involve digital communities in shaping research questions and methodologies (Markham & Buchanan, 2012). Future research could explore alternative models of engagement, such as seeking community input on research design or partnering with subreddit moderators to ensure findings are communicated responsibly.

Ultimately, while Reddit's public-facing nature allows for observational research, ethical best practices require ongoing reflection on how digital scholarship interacts with user agency, data governance, and the broader political economy of online platforms. As the debate over

platform data access and AI-driven content analysis continues, digital researchers must navigate an increasingly complex ethical landscape—balancing the need for scholarly inquiry with evolving concerns over user privacy, platform control, and data commodification.

Challenges in Tracking Deleted Posts

Another key limitation in this study was the difficulty of tracking deleted posts. A notable finding was the relatively low number of removed posts compared to the overall dataset, but understanding the exact reasons for these deletions remained challenging. While the dataset allowed for some identification of removed content, the inability to systematically track deletions over time presented methodological hurdles.

Deleted posts may have been removed by moderators, AutoModerator filters, Reddit administrators, or the original user. However, the exact mechanism of deletion was often unclear. The potential violation tag successfully captured some posts that had been flagged before deletion, but this process was far from exhaustive. Posts can be removed at any time, and unless they attract discussion or controversy, their absence may go unnoticed. This created a blind spot in understanding the full extent of moderation actions, as posts that violated guidelines but were swiftly removed left little trace of their existence.

To address this gap, follow-up requests were posted in each subreddit, asking users whose posts had been deleted to share details about their removal experience. However, these efforts were largely unsuccessful. In r/Philadelphia, the post was automatically deleted because the research account lacked the necessary karma points. An appeal was sent to the moderators, who initially requested further details about the study. After providing a summary of the research and a LinkedIn profile for verification, no further response was received. Similarly, in r/Chicago,

the post was restricted due to karma limitations, and a direct request to the moderators went unanswered.

In r/SanFrancisco, the request remained visible for approximately three hours before being removed. During that time, the few responses it received did not provide insight into deleted posts but instead expressed general frustration with the moderation team. Finally, in r/Austin, the post was deleted almost immediately. Moderators acknowledged the irony of this removal but ultimately cited a rule prohibiting survey requests. However, they did direct the researcher to a repository of deleted posts maintained by the moderation team. Upon reviewing this archive, it became clear that most deletions were due to formatting issues rather than content violations. These challenges highlight the difficulty of obtaining user perspectives on deleted content and the opacity of moderation processes, further emphasizing the need for alternative methods—such as direct interviews with moderators or partnerships with subreddit administrators—to gain deeper insight into content removal practices.

These challenges highlight the opacity of moderation processes and the difficulty of obtaining user perspectives on deleted content. Future research could explore alternative methods, such as direct interviews with moderators or data partnerships with subreddit administrators, to gain a more comprehensive view of content removal practices.

Researcher positionality and influence

Finally, my own experiences with Reddit inevitably shaped the study's design and interpretation. While this study aimed for rigorous objectivity, it is important to acknowledge that research is always conducted from a particular positionality (Rose, 1997). Researchers do not operate as detached observers; rather, they bring their own experiences, assumptions, and biases into their work, which in turn shape the framing of research questions, analytical choices,

and interpretations of data. I first began using Reddit four years ago upon moving to Astoria, Queens. It was recommended to me that I join r/Astoria for local recommendations, landlord reviews, and insights into neighborhood culture. This experience demonstrated how location-based subreddits function as valuable community resources, shaping offline behavior and social integration. A similar reliance on r/Athens occurred upon moving back to Georgia for graduate school, further reinforcing the role of Reddit as a place-based digital infrastructure that extends beyond simple online discussion. These experiences shaped both my perception of Reddit as a community-oriented, participatory space and of what was expected of me as a community member in these places.

In digital research, scholars emphasize the importance of reflexivity—critically examining one’s own role in shaping knowledge production (Markham & Buchanan, 2012). My prior familiarity with Reddit likely influenced expectations about the role of subreddit culture, the balance between moderation and user agency, and the formation of digital community norms. While I attempted to approach the study systematically, it is impossible to fully separate personal engagement with the platform from the analytical lens applied to this research. As Hammersley and Atkinson (2019) argue, "data are always produced through interaction between researcher and researched". Even in observational studies, the researcher’s own subjectivity is inevitably embedded in the process.

This study aligns with broader debates about researcher positionality in digital ethnography, particularly regarding the extent to which scholars can or should maintain distance from the online communities they study. Scholars such as boyd (2014) and Marwick (2013) argue that digital researchers must embrace their dual roles as both participants and analysts, acknowledging that complete detachment is neither possible nor desirable when studying lived

experiences in online spaces. While I approached this study with the intent of capturing an accurate representation of subreddit governance and identity formation, my interpretations were likely influenced by my own experiences navigating digital community spaces.

Recognizing positionality is not a limitation but a necessary aspect of ethical and transparent research (Reich, 2021). By acknowledging the embedded nature of digital ethnography, this study contributes to a growing discourse on how researchers negotiate personal experience, insider knowledge, and analytical objectivity in online research (Orgad, 2009).

Further Study

These limitations do not diminish the study's findings but rather provide critical context for interpreting its scope and implications. The constraints imposed by API access, user resistance to data collection, the challenge of tracking deletions, and the researcher's positionality all underscore the complexities of studying online communities in an era of increasing digital privatization and governance. Future research should explore alternative methods of longitudinal tracking, deeper engagement with subreddit moderators, and more explicit user-informed consent frameworks to address these challenges. Future research should also be conducted by all types of Reddit users— from newcomers to experienced moderators. These varying perspectives add important nuance to the context of digital studies, particularly from an ethnographic standpoint.

One of the most pressing limitations in studying moderation on Reddit is the inability to systematically track deleted content over time. While this study was able to infer some patterns of post removals through discrepancies in dataset retrieval and limited user feedback, a more structured approach is necessary to fully understand the decision-making processes behind deletions. Future studies could implement longitudinal tracking mechanisms, such as automated,

periodic data pulls that capture post statuses at different time intervals. This would provide a clearer picture of when posts are removed and whether they are reinstated. Additionally, incorporating qualitative methods, such as case studies on specific high-profile deletions or real-time tracking of moderation disputes, could offer deeper insight into how content is policed within these communities.

Another critical area for further research is direct engagement with subreddit moderators to gain a more comprehensive understanding of their decision-making processes. While some subreddit moderation teams maintain transparency through public moderation logs or archives, many do not, leaving gaps in understanding how subreddit policies are interpreted and enforced. Future studies could attempt to collaborate with moderators through structured interviews, surveys, or ethnographic participation as a moderator to better contextualize the motivations behind content regulation. Additionally, research into how moderators perceive their roles—whether they see themselves as neutral enforcers, community stewards, or gatekeepers of discourse—could shed light on the power dynamics at play in these spaces. Cross-subreddit comparisons of moderator philosophies and enforcement practices would further illustrate how community identity influences governance.

More explicit user-informed consent frameworks also warrant exploration, particularly in light of ethical concerns regarding data privacy and the use of publicly available posts for academic research. While Reddit's API allows for the collection of public data, the tension between platform policies and user expectations of privacy remains a significant ethical challenge. Many Reddit users express discomfort with their content being extracted for research purposes, particularly without their explicit awareness or consent. Future studies should consider how community-driven consent models such as subreddit opt-in policies for research

participation or real-time user notifications about data usage—could help bridge this gap. This would not only address ethical concerns but also improve the reliability of findings by fostering greater trust between researchers and online communities.

Beyond these methodological considerations, future research should also expand the scope of inquiry into how moderation interacts with broader social and political dynamics. While this study focused on location-based subreddits, similar mechanisms of governance, content control, and community identity formation are at play in political subreddits, hobby-based forums, and identity-driven online spaces. Comparative studies between location-based subreddits and subreddits focused on political discourse, activism, or niche interests could reveal whether these governance structures are unique to geographically oriented communities or part of a larger trend in digital community moderation. Additionally, as Reddit continues to implement new AI-driven moderation tools, future research should examine how automated moderation impacts community trust, discourse quality, and user engagement.

Lastly, as Reddit's role in civic engagement and public discourse grows, future research should investigate its impact on real-world social and political behaviors. How do discussions in location-based subreddits influence local activism, policy discussions, or civic participation? Are these platforms effective at fostering informed debate, or do they reinforce existing social and political divides? Exploring these questions through a combination of qualitative discourse analysis, surveys, and in-depth interviews with active subreddit users could provide crucial insight into the platform's role in shaping contemporary digital and physical communities. By addressing these gaps, future research can build upon the findings of this study to develop a more nuanced understanding of digital governance, community identity, and the evolving relationship between online discourse and offline realities.

REFERENCES

- Aliapoulios, M., Bevensee, E., Blackburn, J., Bradlyn, B., Cristofaro, E. D., Stringhini, G., & Zannettou, S. (2021). *An Early Look at the Parler Online Social Network* (No. arXiv:2101.03820). arXiv. <https://doi.org/10.48550/arXiv.2101.03820>
- Alsaleh, A. (2024). The impact of technological advancement on culture and society. *Scientific Reports*, 14(1), 32140. <https://doi.org/10.1038/s41598-024-83995-z>
- Altman-Lupu, M., & Swanton, B. (2025). *An analysis of Decree 147 on social media*. Project 88.
- Amini, A., Bayiz, Y. E., Ram, A., Marculescu, R., & Topcu, U. (2024). *News Source Credibility Assessment: A Reddit Case Study* (No. arXiv:2402.10938). arXiv. <http://arxiv.org/abs/2402.10938>
- An open letter to Stench: R/Austin*. (2024). https://www.reddit.com/r/Austin/comments/1h9chfx/an_open_letter_to_stench/
- Anand, A., & Pathak, J. (2021). *WallStreetBets against wall street: The Role of reddit in the GameStop short squeeze*. <http://repository.iimb.ac.in/handle/2074/20101>
- Ananthasubramaniam, A., Chen, H., Yan, J., Alkiek, K., Pei, J., Seth, A., Dunagan, L., Choi, M., Litterer, B., & Jurgens, D. (2023). *Exploring Linguistic Style Matching in Online Communities: The Role of Social Context and Conversation Dynamics* (No. arXiv:2307.02758). arXiv. <https://doi.org/10.48550/arXiv.2307.02758>
- Anderson, B. R. O. (1991). *Imagined communities: Reflections on the origin and spread of nationalism* (Revised and extended edition.). Verso.
- Androutsopoulos, J. (2013). *Networked multilingualism: Some language practices on Facebook and their implications*. <https://doi.org/10.1177/1367006913489198>

- Antonucci, T. C., Ajrouch, K. J., & Manalel, J. A. (2017). Social Relations and Technology: Continuity, Context, and Change. *Innovation in Aging*, 1(3), igx029.
<https://doi.org/10.1093/geroni/igx029>
- Anyone know why a police helicopter would just do a lap around the city in the middle of the night?* (2024, December 19). [Reddit Post]. R/Austin.
www.reddit.com/r/Austin/comments/1hi1h8q/anyone_know_why_a_police_helicopter_would_just_do/
- Arsht, A., & Etcovitch, D. (2018, March 2). The Human Cost of Online Content Moderation. *Harvard Journal of Law & Technology*. <https://jolt.law.harvard.edu/digest/the-human-cost-of-online-content-moderation>
- Article19. (2023). *Content moderation and freedom of expression handbook*. UNESCO.
<https://www.article19.org/wp-content/uploads/2023/08/SM4P-Content-moderation-handbook-9-Aug-final.pdf>
- Arttime, O., d'Andrea, V., Gallotti, R., Sacco, P. L., & De Domenico, M. (2020). Effectiveness of dismantling strategies on moderated vs. Unmoderated online social platforms. *Scientific Reports*, 10(1), 14392. <https://doi.org/10.1038/s41598-020-71231-3>
- Atton, C., & Hamilton, J. F. (2008). *Alternative journalism*. Sage.
- Babar, Y. (2020). *A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF MINNESOTA*.
- Barasa, S. N. (2013). *Leetspeak in linguistic taboos*.
- Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American Society for Information Science and Technology*, 59(9), 1493–1512. <https://doi.org/10.1002/asi.20857>

- Bellah, R. N. (1995). *Community Properly Understood: A Defense of “Democratic Communitarianism.”*
- Benkler, Y. (2007). The Wealth of Networks: How Social Production Transforms Markets and Freedom. *Prometheus*, 25(1). <https://doi.org/10.1080/08109020701191257>
- Benson, R. (2009). Shaping the Public Sphere: Habermas and Beyond. *The American Sociologist*, 40(3), 175–197. <https://doi.org/10.1007/s12108-009-9071-4>
- Best place to buy a Christmas tree?* (2024, November 29). [Reddit Post]. R/Sanfrancisco. www.reddit.com/r/sanfrancisco/comments/1h2ptvq/best_place_to_buy_a_christmas_tree/
- Birks, M., & Mills, J. (2015). *Grounded Theory: A Practical Guide*. SAGE.
- Boellstorff, T. (with Nardi, B. A., Pearce, C., Taylor, T. L., & Marcus, G.). (2012). *Ethnography and virtual worlds: A handbook of method*. University Press.
- Boler, M. (2008). *Digital media and democracy: Tactics in hard times*. MIT Press.
- Bourdieu, P. (1990). *The Logic of Practice*. Stanford University Press.
- Bourdieu, P. (1991). *Language and Symbolic Power*. Harvard University Press.
- boyd, danah. (2010). Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. In Z. Papacharissi (Ed.), *A Networked Self* (0 ed., pp. 47–66). Routledge. <https://doi.org/10.4324/9780203876527-8>
- Boyd, D. (2014). *It’s complicated: The social lives of networked teens*. Yale University Press.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic Analysis. In P. Liamputtong (Ed.), *Handbook of Research Methods in Health Social Sciences* (pp. 843–860). Springer Nature. doi.org/10.1007/978-981-10-5251-4_103

- Brewer, M. B., & Silver, M. D. (2000). Group distinctiveness, social identification, and collective mobilization. In *Self, identity, and social movements*. (pp. 153–171). University of Minnesota Press. <https://research.ebsco.com/linkprocessor/plink?id=32ce79ea-d97e-3dc8-93d7-4784fe9eec65>
- Bruns, A. (2011). Gatekeeping, gatewatching, real-time feedback. *Brazilian Journalism Research*, 7(11).
- Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, 7(4–5), 585–614. <https://doi.org/10.1177/1461445605054407>
- Buckingham, S. L., Brodsky, A. E., Rochira, A., Fedi, A., Mannarini, T., Emery, L., Godsay, S., Miglietta, A., & Gattino, S. (2018). Shared Communities: A Multinational Qualitative Study of Immigrant and Receiving Community Members. *American Journal of Community Psychology*, 62(1–2), 23–40. <https://doi.org/10.1002/ajcp.12255>
- Caliandro, A. (2018). Digital Methods for Ethnography: Analytical Concepts for Ethnographers Exploring Social Media Environments. *Journal of Contemporary Ethnography*, 47(5), 551–578. <https://doi.org/10.1177/0891241617702960>
- Caplan, R. (2018). *Content or Context Moderation*. Data & Society. https://datasociety.net/wp-content/uploads/2018/11/DS_Content_or_Context_Moderation.pdf
- Carey, J. W. (1975). *A cultural approach to communication*.
- Carey, J. W. (1983). Technology and Ideology: The Case of the Telegraph. *Prospects*, 8, 303–325. <https://doi.org/10.1017/S0361233300003793>
- Carey, J. W. (2005). Historical pragmatism and the internet. *New Media & Society*, 7(4), 443–455. <https://doi.org/10.1177/1461444805054107>

Ceci, L. (2024). *U.S. number of logged-in Reddit users 2025*. Statista.

<https://www.statista.com/statistics/1018997/number-of-reddit-users-usa/>

Cedar fever hack? (2024, December 15). [Reddit Post]. R/Austin.

www.reddit.com/r/Austin/comments/1hera3u/cedar_fever_hack/

Centivany, A. (2016). Values, ethics and participatory policymaking in online communities.

Proceedings of the Association for Information Science and Technology, 53(1), 1–10.

<https://doi.org/10.1002/pra2.2016.14505301058>

Chan, K., Ortutay, B., & Riccardi, N. (2025, January 7). Meta eliminates fact-checking in latest

bow to Trump | AP News. *AP News*. <https://apnews.com/article/meta-facts-trump-musk-community-notes-413b8495939a058ff2d25fd23f2e0f43>

Chandrasekharan, E., Samory, M., Jhaver, S., Charvat, H., Bruckman, A., Lampe, C., Eisenstein,

J., & Gilbert, E. (2018). The Internet's Hidden Rules: An Empirical Study of Reddit

Norm Violations at Micro, Meso, and Macro Scales. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1–25. <https://doi.org/10.1145/3274301>

Chayko, M. (2002). *Connecting—How we Form Social Bonds and Communities in the Internet*

Age. SUNY Press. <https://www.thetedkarchive.com/library/mary-chayko-connecting>

Chowdhury, N., & Keller, D. (2022). *Automated Content Moderation: A Primer*. [https://fsi-](https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/automated_content_moderation_a_primer.pdf)

[live.s3.us-west-1.amazonaws.com/s3fs-](https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/automated_content_moderation_a_primer.pdf)

[public/automated_content_moderation_a_primer.pdf](https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/automated_content_moderation_a_primer.pdf)

Coeckelbergh, M., & Reader, J. (2023). Searching for the Good Life: Boundary Crossing, Co-

Existence, and Conviviality. *Postdigital Science and Education*, 1–13.

<https://doi.org/10.1007/s42438-023-00406-8>

Communications Decency Act, § 230 47 U.S.C. (1996).

<https://www.govinfo.gov/content/pkg/USCODE-2020-title47/pdf/USCODE-2020-title47-chap5-subchapII-partI-sec230.pdf>

Coutinho, P., & José, R. (2017). Moderation techniques for user-generated content in place-based communication. *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–6. <https://doi.org/10.23919/CISTI.2017.7975786>

Cowgill, B., & Tucker, C. (2019). *Algorithmic Fairness and Economics*.

Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., & Potts, C. (2013). No country for old members: User lifecycle and linguistic change in online communities. *Proceedings of the 22nd International Conference on World Wide Web*, 307–318. <https://doi.org/10.1145/2488388.2488416>

Danesi, M. (2014). Forging a linguistic identity in the age of the Internet. *Forum Italicum: A Journal of Italian Studies*, 48(2), 227–237. <https://doi.org/10.1177/0014585814529223>

Davidson, T., Warmesley, D., Macy, M., & Weber, I. (2017). Automated Hate Speech Detection and the Problem of Offensive Language. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), Article 1. <https://doi.org/10.1609/icwsm.v11i1.14955>

Davies, C., & Milmo, D. (2025, February 25). Ofcom calls on tech firms to step up action against ‘revenge porn.’ *The Guardian*. <https://www.theguardian.com/society/2025/feb/25/ofcom-calls-on-tech-firms-to-step-up-action-against-revenge-porn>

Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2023, H.R. 5586, 118th Congress 1st Session (2023). <https://www.congress.gov/118/bills/hr5586/BILLS-118hr5586ih.pdf>

- Demony, C. (2025, January 7). Britain to make sexually explicit “deepfakes” a crime. *Reuters*.
<https://www.reuters.com/world/uk/britain-make-sexually-explicit-deepfakes-crime-2025-01-07/>
- Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, 51(3), 629–636. <https://doi.org/10.1037/h0046408>
- Dijck, J. van, Poell, T., & Waal, M. de. (2018). *The Platform Society: Public Values in a Connective World*. Oxford University Press.
- Dixon, S. J. (2024). *U.S. market share of leading social media websites 2024*. Statista.
<https://www.statista.com/statistics/265773/market-share-of-the-most-popular-social-media-websites-in-the-us/>
- Dotson, T. (2017). *Technically Together: Reconstructing Community in a Networked World*. MIT Press.
- Dotson, T. C. (2015). *Communitarian Technology: Realizing Thicker Communities Within Societies Built for Networked Individualism*.
- Dowling, M.-E. (2023). Far-right populism in alt-tech: A challenge for democracy? *New Media & Society*, 14614448231205889. <https://doi.org/10.1177/14614448231205889>
- Driskell, R. B., & Lyon, L. (2002). Are Virtual Communities True Communities? Examining the Environments and Elements of Community. *City & Community*, 1(4), 373–390.
<https://doi.org/10.1111/1540-6040.00031>
- Dwork, C., Hays, C., Kleinberg, J., & Raghavan, M. (2023). *Content Moderation and the Formation of Online Communities: A Theoretical Framework* (No. arXiv:2310.10573). arXiv. <http://arxiv.org/abs/2310.10573>

Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, 41(Volume 41, 2012), 87–100. <https://doi.org/10.1146/annurev-anthro-092611-145828>

Eckert, P. (2018). *Meaning and linguistic variation: The third wave in sociolinguistics* / Penelope Eckert, Stanford University, California.
<https://research.ebsco.com/linkprocessor/plink?id=08a9fafa-27fa-3579-8e8b-5db11fa569f5>

Fairmount Hardware Christmas trees? (2024, November 27). [Reddit Post]. R/Philadelphia.
www.reddit.com/r/philadelphia/comments/1h19jo0/fairmount_hardware_christmas_trees/

Feenberg, A., & Barney, D. D. (2004). *Community in the digital age: Philosophy and practice*. Rowman & Littlefield.
<https://books.google.com/books?hl=en&lr=&id=OhGZD9ItPKEC&oi=fnd&pg=PP9&dq=related:1TSWyTAsZBIJ:scholar.google.com/&ots=EdngyuSz1p&sig=WCGyv10QV1U6YWxPEuaiIkpZv84>

Fibrianto, A. S., & Yuniar, A. D. (2019). Technological Development and its Impact on Community Social Behavior. *Proceedings of the 1st International Conference on Social Knowledge Sciences and Education (ICSKSE 2018)*. Proceedings of the 1st International Conference on Social Knowledge Sciences and Education (ICSKSE 2018), Batu, Indonesia. <https://doi.org/10.2991/icskse-18.2019.42>

Fiesler, C., Jiang, J., McCann, J., Frye, K., & Brubaker, J. (2018). Reddit Rules! Characterizing an Ecosystem of Governance. *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1), Article 1. <https://doi.org/10.1609/icwsm.v12i1.15033>

- Forestal, J. (2022). *Designing for Democracy: How to Build Community in Digital Environments*. Oxford University Press.
- Franco, A., & Birenboim, A. (2024). The Interrelations Between Virtual and Physical Spaces: The Case of Smartphone Usage Among Adolescents. *Annals of the American Association of Geographers*, 114(9), 1948–1967. <https://doi.org/10.1080/24694452.2024.2367675>
- franzke, aline shakti, Bechmann, A., Zimmer, M., & Ess, C. M. (2019). *Internet Research: Ethical Guidelines 3.0*. <https://aoir.org/reports/ethics3.pdf>
- Gallagher, S. E., & Savage, T. (2015). “What is, Becomes What is Right”: A Conceptual Framework of Newcomer Legitimacy for Online Discussion Communities. *Journal of Computer-Mediated Communication*, 20(4), 400–416. <https://doi.org/10.1111/jcc4.12122>
- Gergen, K. J. (2003). *Self and Community in the New Floating Worlds*.
- Gil De Zúñiga, H., Jung, N., & Valenzuela, S. (2012). Social Media Use for News and Individuals’ Social Capital, Civic Engagement and Political Participation. *Journal of Computer-Mediated Communication*, 17(3), 319–336. <https://doi.org/10.1111/j.1083-6101.2012.01574.x>
- Gill, R. (2023, June 20). *Nudging users to report potentially harmful online content*. [Www.Ofcom.Org.Uk. https://www.ofcom.org.uk/online-safety/safety-technology/nudging-users-to-report-potentially-harmful-online-content/](https://www.ofcom.org.uk/online-safety/safety-technology/nudging-users-to-report-potentially-harmful-online-content/)
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.
- Glaser, B. G. (1965). The Constant Comparative Method of Qualitative Analysis. *Social Problems*, 12(4), 436–445.

- Gongane, V. U., Munot, M. V., & Anuse, A. D. (2022). Detection and moderation of detrimental content on social media platforms: Current status and future directions. *Social Network Analysis and Mining*, 12(1), 129. <https://doi.org/10.1007/s13278-022-00951-3>
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). *Algorithmic content moderation: Technical and political challenges in the automation of platform governance*. <https://doi.org/10.1177/2053951719897945>
- Graham, M. (2024, June 26). *Deepfakes: Federal and state regulation aims to curb a growing threat*. Thomson Reuters Institute. <https://www.thomsonreuters.com/en-us/posts/government/deepfakes-federal-state-regulation/>
- Gray, J., Sandvoss, C., & Harrington, C. L. (Eds.). (2017). *Fandom: Identities and communities in a mediated world* (Second Edition). New York University Press.
- Greer, S. (2017). *The Concept of Community: Readings with Interpretations*. Routledge. <https://doi.org/10.4324/9781315131443>
- Gregorio, G. D., & Goanta, C. (2023). *Content Monetization*. https://law.yale.edu/sites/default/files/area/center/isp/documents/contentmonetization_isp_essayseries_2023.pdf
- Greis, M., Alt, F., Henze, N., & Memarovic, N. (2014). I can wait a minute: Uncovering the optimal delay time for pre-moderated user-generated content on public displays. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1435–1438. <https://doi.org/10.1145/2556288.2557186>
- Gruzd, A., Soares, F. B., & Mai, P. (2023). Trust and Safety on Social Media: Understanding the Impact of Anti-Social Behavior and Misinformation on Content Moderation and Platform

- Governance. *Social Media + Society*, 9(3), 20563051231196878.
<https://doi.org/10.1177/20563051231196878>
- Hammersley, M., & Atkinson, P. (2019). *Ethnography: Principles in Practice* (4th ed.). Routledge. <https://doi.org/10.4324/9781315146027>
- Hampton, K. N. (2004). Networked sociability online, off-line. *The Network Society: A Cross-Cultural Perspective*, 217–232.
- Hampton, K. N., & Wellman, B. (2001). Long Distance Community in the Network Society: Contact and Support Beyond Netville. *American Behavioral Scientist*, 45(3), 476–495.
<https://doi.org/10.1177/00027640121957303>
- Hargittai, E. (2022). *Connected in Isolation: Digital Privilege in Unsettled Times*. MIT Press.
- Hartzell, A., Jackson, B., & Pinchot, J. (2021). Source Credibility in Social Media: A Case Study of a Reddit Community. *Issues In Information Systems*, 22(3), 212–223.
https://doi.org/10.48009/3_iis_2021_230-241
- Heller, M. (2011). *Paths to Post-nationalism: A Critical Ethnography of Language and Identity*. Oxford University Press.
- Hine, C. (2015). *Ethnography for the Internet: Embedded, embodied and everyday* / Christine Hine. <https://research.ebsco.com/linkprocessor/plink?id=11447d13-233f-3d7c-903e-a4bf923faa54>
- Hintz, E. A., & Betts, T. (2022). Reddit in communication research: Current status, future directions and best practices. *Annals of the International Communication Association*, 46(2), 116–133. <https://doi.org/10.1080/23808985.2022.2064325>

- Hogg, M. A. (2010). Influence and leadership. In *Handbook of social psychology, Vol. 2, 5th ed* (pp. 1166–1207). John Wiley & Sons, Inc.
<https://doi.org/10.1002/9780470561119.socpsy002031>
- Horrigan, J. B. (2001). *Online Communities*. Pew Research Center.
<https://www.pewresearch.org/internet/2001/10/31/online-communities/>
- Hosek, A. M., & Austin, L. (2016). Exploring Pedagogical and Emotional Response in Higher Education Classrooms During the Boston Marathon Bombing Crisis. *Qualitative Research Reports in Communication*, 17(1), 68–76.
<https://doi.org/10.1080/17459435.2016.1143386>
- Inserra, D. (2024). *A Guide to Content Moderation for Policymakers*. Cato Institute.
- Jenkins, H. (2013). *Spreadable media: Creating value and meaning in a networked culture* / Henry Jenkins, Sam Ford, & Joshua Green.
<https://research.ebsco.com/linkprocessor/plink?id=b0a5ee2f-a9a5-3f4b-92e1-c5b0e5bce2f2>
- Jhaver, S., Appling, D. S., Gilbert, E., & Bruckman, A. (2019). “Did You Suspect the Post Would be Removed?”: Understanding User Reactions to Content Removals on Reddit. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–33.
<https://doi.org/10.1145/3359294>
- Jiménez-Durán, R. (2023). The Economics of Content Moderation: Theory and Experimental Evidence from Hate Speech on Twitter. *George J. Stigler Center for the Study of the Economy & the State, Working Paper No. 324*. <https://doi.org/10.2139/ssrn.4590147>
- Katz, J. E., & Rice, R. E. (2002). *Social Consequences of Internet Use: Access, Involvement, and Interaction*. MIT Press.

- Katz, J. E., Rice, R. E., Acord, S., Dasgupta, K., & David, K. (2004). Personal Mediated Communication and the Concept of Community in Theory and Practice. *Annals of the International Communication Association*, 28(1), 315–371.
<https://doi.org/10.1080/23808985.2004.11679039>
- Kavanaugh, A. L. (2014). Web Communities Versus Physical Communities. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of Social Network Analysis and Mining* (pp. 2344–2355). Springer. https://doi.org/10.1007/978-1-4614-6170-8_225
- Kavoura, A. (2014). Social media, online imagined communities and communication research. *Library Review*, 63(6/7), 490–504. <https://doi.org/10.1108/LR-06-2014-0076>
- Keller, D. (2022). The DSA’s Industrial Model for Content Moderation. *Verfassungsblog*.
<https://doi.org/10.17176/20220224-121133-0>
- Kidd, J. (2020). Infrastructures that democratize? In P. Hetland, P. Pierroux, & L. Esborg (Eds.), *A History of Participation in Museums and Archives: Traversing Citizen Science and Citizen Humanities* (1st ed.). Routledge. <https://doi.org/10.4324/9780429197536>
- Killion, V. L. (2024). *Freedom of Speech: An Overview* (No. R47986). Congressional Research Service. <https://crsreports.congress.gov>
- Klonick, K. (2018). THE NEW GOVERNORS: THE PEOPLE, RULES, AND PROCESSES GOVERNING ONLINE SPEECH. *HARVARD LAW REVIEW*, 131.
- Klug, D., Steen, E., & Yurechko, K. (2023). How Algorithm Awareness Impacts Algospeak Use on TikTok. *Companion Proceedings of the ACM Web Conference 2023*, 234–237.
<https://doi.org/10.1145/3543873.3587355>

- Koebler, J. (2023, June 14). *The Reddit Protest Is a Battle for the Soul of the Human Internet*.
https://www.reddit.com/r/ModCoord/comments/148ks6u/indefinite_blackout_next_steps_polling_your/
- Kolozsvari, O. (2012). *Constructing Time and Space and Transcending Boundaries in Long-Distance Relationships* [Georgia State University]. <https://doi.org/10.57709/3074378>
- LaFrance, A. (2025, March 4). The Nicest Swamp on the Internet. *The Atlantic*.
<https://www.theatlantic.com/magazine/archive/2025/04/reddit-culture-community-credibility/681765/>
- Lane, J., & Lingel, J. (2022). Digital Ethnography for Sociology: Craft, Rigor, and Creativity. *Qualitative Sociology*, 45(3), 319–326. <https://doi.org/10.1007/s11133-022-09509-3>
- Langvardt, K. (2018). *Regulating Online Content Moderation*. 106.
- Lauer, S., Wong, K. L. Y., & Yan, M. C. (2024). Social infrastructure, community organizations, and friendship formation: A scoping review. *Community Development Journal*, bsae023.
<https://doi.org/10.1093/cdj/bsae023>
- Lee, E. (2020). Moderating Content Moderation: A Framework for Nonpartisanship in Online Governance. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3705466>
- Leibmann, L., Weld, G., Zhang, A. X., & Althoff, T. (2025). *Reddit Rules and Rulers: Quantifying the Link Between Rules and Perceptions of Governance across Thousands of Communities* (No. arXiv:2501.14163). arXiv. <https://doi.org/10.48550/arXiv.2501.14163>
- Leuckert, S., & Leuckert, M. (2020). Towards a digital sociolinguistics: Communities of Practice on Reddit. In S. Rüdiger & D. Dayter (Eds.), *Studies in Corpus Linguistics* (Vol. 98, pp. 15–40). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.98.01leu>
- Lingel, J. (2017). *Digital countercultures and the struggle for community*. MIT Press.

- Lingel, J. (2021). *The Gentrification of the Internet: How to Reclaim Our Digital Freedom* (1st ed.). University of California Press. <https://doi.org/10.2307/j.ctv1h9dkf9>
- Lingel, J., Trammell, A., Sanchez, J., & Naaman, M. (2012). Practices of information and secrecy in a punk rock subculture. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 157–166.
<https://doi.org/10.1145/2145204.2145230>
- Liu, Y., Yildirim, P., & Zhang, Z. J. (2022). Implications of Revenue Models and Technology for Content Moderation Strategies. *Marketing Science*, 41(4), 831–847.
<https://doi.org/10.1287/mksc.2022.1361>
- Live Christmas Trees in South/Central Austin*. (2024, December 10). [Reddit Post]. R/Austin.
www.reddit.com/r/Austin/comments/1hba27i/live_christmas_trees_in_southcentral_austin/
- Lorenz, T. (2022, April 12). Internet “algospeak” is changing our language in real time, from “nip nops” to “le dollar bean”. *The Washington Post*.
- Lüders, A., Dinkelberg, A., & Quayle, M. (2022). Becoming “us” in digital spaces: How online users creatively and strategically exploit social media affordances to build up social identity. *Acta Psychologica*, 228, 103643. <https://doi.org/10.1016/j.actpsy.2022.103643>
- Markham, A. (2021). The limits of the imaginary: Challenges to intervening in future speculations of memory, data, and algorithms. *New Media & Society*, 23(2), 382–405.
<https://doi.org/10.1177/1461444820929322>
- Markham, A., & Buchanan, E. (2012). *Recommendations from the AoIR Ethics Working Committee (Version 2.0)*.

- Marwick, A., Clancy, B., & Furl, K. (2022). Far-Right Online Radicalization: A Review of the Literature. *The Bulletin of Technology & Public Life*.
<https://doi.org/10.21428/bfcb0bff.e9492a11>
- Marwick, A. E. (2013). *Status Update: Celebrity, Publicity, and Branding in the Social Media Age*. Yale University Press.
<http://ebookcentral.proquest.com/lib/ugalib/detail.action?docID=3421330>
- Massanari, A. L. (2015). *Participatory Culture, Community, and Play*. Peter Lang US.
<https://doi.org/10.3726/978-1-4539-1501-1>
- Matias, J. N. (2019a). Preventing harassment and increasing group participation through social norms in 2,190 online science discussions. *Proceedings of the National Academy of Sciences*, 116(20), 9785–9789. <https://doi.org/10.1073/pnas.1813486116>
- Matias, J. N. (2019b). The Civic Labor of Volunteer Moderators Online. *Social Media + Society*.
<https://doi.org/10.1177/2056305119836778>
- Mazzurco, S. (2023). Content Moderation Regulation as Legal Role-Scripting. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4401151>
- McDonald, K. (2015). From Indymedia to Anonymous: Rethinking action and identity in digital cultures. *Information, Communication & Society*, 18(8), 968–982.
<https://doi.org/10.1080/1369118X.2015.1039561>
- Mendelsohn, J., Vijan, M., Card, D., & Budak, C. (2024). Framing Social Movements on Social Media: Unpacking Diagnostic, Prognostic, and Motivational Strategies. *Journal of Quantitative Description: Digital Media*, 4. <https://doi.org/10.51685/jqd.2024.icwsm.9>

- Mensah, H., Xiao, L., & Soundarajan, S. (2020). Characterizing the Evolution of Communities on Reddit. *International Conference on Social Media and Society*, 58–64.
<https://doi.org/10.1145/3400806.3400814>
- Mercea, D. (2016). *Civic Participation in Contentious Politics: The Digital Foreshadowing of Protest*. Springer.
- Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, 59, 210–220.
<https://doi.org/10.1016/j.pragma.2013.07.012>
- Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & Mccann, R. M. (2003). Credibility for the 21st Century: Integrating Perspectives on Source, Message, and Media Credibility in the Contemporary Media Environment. *Annals of the International Communication Association*, 27(1), Article 1. <https://doi.org/10.1080/23808985.2003.11679029>
- Meyrowitz, J. (1985). *No sense of place: The impact of electronic media on social behavior / Joshua Meyrowitz*. (1st ed.). Oxford University Press.
<https://research.ebsco.com/linkprocessor/plink?id=f37f8dc0-1c8a-3d73-b3ba-b250d79d115f>
- Miller, C. (2015). *Life in the New Media landscape: 2*.
- Miller, R. A. (2017). “My Voice Is Definitely Strongest in Online Communities”: Students Using Social Media for Queer and Disability Identity-Making. *Journal of College Student Development*, 58(4), 509–525. <https://doi.org/10.1353/csd.2017.0040>
- Milner, R. M. (2018). *The World Made Meme: Public Conversations and Participatory Media*. MIT Press.

- Mods removed my post about Nancy Pelosi shafting AOC saying it “wasn’t relevant.”* (2024, December 17). [Reddit Post]. R/Sanfrancisco.
www.reddit.com/r/sanfrancisco/comments/1hgnssg/mods_removed_my_post_about_nancy_pelosi_shafting/
- Myers West, S. (2018). Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society*, 20(11), 4366–4383.
<https://doi.org/10.1177/1461444818773059>
- Naqvi, F. A. (2024, July 1). Reddit Tightens Grip on Data with AI Protection Plan. *DigiAlps LTD*. <https://digiалps.com/reddit-tightens-grip-on-data-with-ai-protection-plan/>
- Nardi, B. A. (2010). *My life as a night elf priest: An anthropological account of World of warcraft*. University of Michigan Press.
- Nardi, B. A. (2016). When fieldnotes seem to write themselves. In S. W. Tratner & R. Sanjek (Eds.), *eFieldnotes: The makings of anthropology in the digital world* (pp. 192–210). University of Pennsylvania Press.
- New policy: Public officials’ posts can no longer be taken down by user flagging.* (2024, November 21). [Reddit Post]. R/Sanfrancisco.
www.reddit.com/r/sanfrancisco/comments/1gwjn2i/new_policy_public_officials_posts_can_no_longer/
- Nicas, J., & Alba, D. (2021, January 10). Amazon, Apple and Google Cut Off Parler, an App That Drew Trump Supporters. *The New York Times*.
<https://www.nytimes.com/2021/01/09/technology/apple-google-parler.html>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism—University of Georgia*. <https://galileo->

uga.primo.exlibrisgroup.com/discovery/fulldisplay?context=L&vid=01GALI_UGA:UGA&search_scope=MyInstitution&tab=LibraryCatalog&docid=alma9949105971302959

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, 16(1), 1609406917733847. <https://doi.org/10.1177/1609406917733847>

Nurture Originals, Foster Art, and Keep Entertainment Safe Act of 2024, Senate 118th Congress 2d Session, S. 4875 (2024).

Nurture Originals, Foster Art, and Keep Entertainment Safe Act of 2024, H. R. 9551, 118th Congress 2nd Session (2024). <https://www.congress.gov/118/bills/hr9551/BILLS-118hr9551ih.pdf>

Ocean Avenue 24hr Fitness Incident. (2024, November 20). [Reddit Post]. R/Sanfrancisco. www.reddit.com/r/sanfrancisco/comments/1gvdj9t/ocean_avenue_24hr_fitness_incident/

Oddný, L., Ainslie, C., Lakshman, S., & Nathan, D. (2023). Impact of Reddit Community Culture on User Attitude Expression and Social Interaction. *Journal of Linguistics and Communication Studies*, 2(4), 61–67. <https://doi.org/10.56397/JLCS.2023.12.07>

Oksanen, A., Celuch, M., Oksa, R., & Savolainen, I. (2024). Online communities come with real-world consequences for individuals and societies. *Communications Psychology*, 2(1), 71. <https://doi.org/10.1038/s44271-024-00112-6>

Orgad, S. (2009). How Can Researchers Make Sense of the Issues Involved in Collecting and Interpreting Online and Offline Data? In A. Markham & N. Baym, *Internet Inquiry: Conversations About Method* (pp. 33–53). SAGE Publications, Inc. <https://doi.org/10.4135/9781483329086.n4>

- Page, R. (2018). *Narratives Online: Shared Stories in Social Media*. Cambridge University Press.
- Papacharissi, Z. (Ed.). (2010). Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. In *A Networked Self* (0 ed., pp. 47–66). Routledge.
<https://doi.org/10.4324/9780203876527-8>
- Papacharissi, Z. (2015). *Affective Publics: Sentiment, Technology, and Politics*. Oxford University Press.
- Park, J. S., Seering, J., & Bernstein, M. S. (2022). *Measuring the Prevalence of Anti-Social Behavior in Online Communities* (No. arXiv:2208.13094). arXiv.
<https://doi.org/10.48550/arXiv.2208.13094>
- Park, R. E. (1936). Human Ecology. *American Journal of Sociology*, 42(1), 1–15.
- People who don't wear scarves—You're really missing out.* (2024, December 12). [Reddit Post]. R/Chicago.
www.reddit.com/r/chicago/comments/1hcn2qi/people_who_dont_wear_scarves_youre_really_missing/
- Peters, G. (2022). *The Rising Threat of Domestic Terrorism: A Review of the Federal Response to Domestic Terrorism and the Spread of Extremist Content on Social Media*. Homeland Security & Governmental Affairs. https://www.hsgac.senate.gov/wp-content/uploads/imo/media/doc/221116_HSGACMajorityReport_DomesticTerrorism%26SocialMedia.pdf
- Peters, J., & Johnson, B. (2016). Conceptualizing private governance in a networked society. *North Carolina Journal of Law & Technology*, 18(1).

Peters, J., & Sato, M. (2023, January 16). Here's the note Reddit sent to moderators threatening them if they don't reopen | The Verge. *The Verge*.

<https://www.theverge.com/2023/6/16/23763538/reddit-blackout-api-protest-mod-replacement-threat>

Petrosyan, A. (2025). *Internet and social media users in the world 2025*. Statista.

<https://www.statista.com/statistics/617136/digital-population-worldwide/>

Phillips. (2019). *It Wasn't Just the Trolls: Early Internet Culture, "Fun," and the Fires of Exclusionary Laughter*. <https://doi.org/10.1177/2056305119849493>

Phillipson, R. (2018). Language challenges in global and regional integration. *Darnioji Daugiakalbystė*, 12, 14–35.

Piller, I. (2016). *Linguistic Diversity and Social Justice: An Introduction to Applied Sociolinguistics*. Oxford University Press.

Pink, S. (2016). *Digital ethnography: Principles and practice* / Sarah Pink, Heather Horst, John Postill, Larissa Hjorth, Tania Lewis, Jo Tacchi.

<https://research.ebsco.com/linkprocessor/plink?id=f05bd851-4889-3215-9e7e-8f569026369a>

Police helicopter circling Bouldin. (2024, December 17). [Reddit Post]. R/Austin.

www.reddit.com/r/Austin/comments/1hg3lo9/police_helicopter_circling_bouldin/

Postigo, H. (2003). Emerging Sources of Labor on the Internet: The Case of America Online Volunteers. *International Review of Social History*, 48(S11), 205–223.

<https://doi.org/10.1017/S0020859003001329>

Potts, L., Small, R., & Trice, M. (2019). Boycotting the Knowledge Makers: How Reddit Demonstrates the Rise of Media Blacklists and Source Rejection in Online Communities.

- IEEE Transactions on Professional Communication*, 62(4), 351–363. IEEE Transactions on Professional Communication. <https://doi.org/10.1109/TPC.2019.2946942>
- Preece, J. (2000). *Online communities: Designing usability, supporting sociability* / Jenny Preece. John Wiley. <https://research.ebsco.com/linkprocessor/plink?id=65abb4d4-1dd0-32d7-a8d7-91727414a8ba>
- Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. *Social Media + Society*, 7(2), 205630512110190. <https://doi.org/10.1177/20563051211019004>
- Purcell, P. A. (Ed.). (2006). *Networked neighbourhoods: The connected community in context*. Springer.
- Putnam, R. D. (2001). *Bowling alone: The collapse and revival of American community* (1st ed.). Simon & Schuster.
- R/Austin FAQ*. (n.d.). Reddit. Retrieved March 5, 2025, from <https://www.reddit.com/r/Austin/wiki/faq/>
- R/Austin Rules and Moderation*. (n.d.). Reddit. Retrieved March 5, 2025, from <https://www.reddit.com/r/Austin/wiki/rulesandmoderating/>
- R/Chicago*. (n.d.). Reddit. Retrieved March 5, 2025, from <https://www.reddit.com/r/chicago/>
- Reddiquette*. (2025, January 28). Reddit Help. <https://support.reddithelp.com/hc/en-us/articles/205926439-Reddiquette>
- Reddit. (2025, February 14). *Developer Platform & Accessing Reddit Data*. Reddit Help. <https://support.reddithelp.com/hc/en-us/articles/14945211791892-Developer-Platform-Accessing-Reddit-Data>

- Reddit*—<https://i.redd.it/bemzhsgcz97e1.jpeg>. (n.d.). [Online post]. Retrieved March 5, 2025, from <https://www.reddit.com/media?url=https%3A%2F%2Fi.redd.it%2Fbemzhsgcz97e1.jpeg>
- Reddy, H., & Chandrasekharan, E. (2023). Evolution of Rules in Reddit Communities. *Computer Supported Cooperative Work and Social Computing*, 278–282. <https://doi.org/10.1145/3584931.3606973>
- Reich, J. A. (2021). Power, Positionality, and the Ethic of Care in Qualitative Research. *Qualitative Sociology*, 44(4), 575–581. <https://doi.org/10.1007/s11133-021-09500-4>
- Renaud, K., & Coles-Kemp, L. (2022). Accessible and Inclusive Cyber Security: A Nuanced and Complex Challenge. *Sn Computer Science*, 3(5), 346. <https://doi.org/10.1007/s42979-022-01239-1>
- Riley, J. K., & Cowart, H. S. (2021). *The Reddit Oasis: Analyzing the potential role of location-based subreddits in the alleviation of news deserts*.
- Rising, D. (2025, February 18). Vietnam’s strict new social media regulations strangle free speech, report says. *AP News*. <https://apnews.com/article/vietnam-social-media-restrictions-flc421efb3e10e6cfb433c01e88a8eda>
- R/MissingPersons*. (n.d.). Reddit. Retrieved March 5, 2025, from https://www.reddit.com/r/MissingPersons/wiki/rules/#wiki_2.29_social_networking_web_sites_are_not_automatically_accepted
- Roberts, S. (2016). *Commercial Content Moderation: Digital Laborers’ Dirty Work*.
- Roberts, S. (2017). *Content moderation*. <https://escholarship.org/content/qt7371c1hf/qt7371c1hf.pdf>
- Roberts, S. (2019). *Behind the Screen*. Yale University Press.

- Rollero, C., & De Piccoli, N. (2010). Does place attachment affect social well-being? *European Review of Applied Psychology*, 60(4), 233–238.
<https://doi.org/10.1016/j.erap.2010.05.001>
- Rose, G. (1997). Situating knowledges: Positionality, reflexivities and other tactics. *Progress in Human Geography*, 21(3), 305–320. <https://doi.org/10.1191/030913297673302122>
- Rosenzweig, P. (2024). *Content Moderation and the Least Cost Avoider*.
- Rozenshtein, A. Z., & Langvardt, K. (2025). Beyond the Editorial Analogy: First Amendment Protections for Platform Content Moderation After *Moody v. NetChoice*. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5131132>
- /r/SanFrancisco town hall: Should public officials' posts be exempt from flagging? (2024, November 20). [Reddit Post]. R/Sanfrancisco.
www.reddit.com/r/sanfrancisco/comments/1gvcart/rsanfrancisco_town_hall_should_public_officials/
- Sallafranque-St-Louis, F., & Normand, C. L. (2017). From solitude to solicitation: How people with intellectual disability or autism spectrum disorder use the internet. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 11(1).
<https://doi.org/10.5817/CP2017-1-7>
- San Francisco. (n.d.). Reddit. Retrieved March 5, 2025, from
<https://www.reddit.com/r/sanfrancisco/>
- Sap, M., Card, D., Gabriel, S., Choi, Y., & Smith, N. A. (2019). The Risk of Racial Bias in Hate Speech Detection. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1668–1678). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1163>

- Scheffer, J. (2024). *Mirrored Spaces: Social Inequality in the Digital Age*. Springer Nature.
- Schmitz, A., & Samory, M. (2025). From Volunteerism to Corporatization: Analyzing Participation in the 2015 and 2023 Reddit Blackouts. *Social Media + Society*, 11(1), 20563051241309497. <https://doi.org/10.1177/20563051241309497>
- Serious Discussion Thread: The Mods are dorks here , care to elaborate anyone?* (2024, December 15). [Reddit Post]. R/Philadelphia.
www.reddit.com/r/philadelphia/comments/1heysiw/serious_discussion_thread_the_mods_are_dorks_here/
- Shannon, C. E., & Weaver, W. (1949). A mathematical model of communication. *Urbana, IL: University of Illinois Press*, 11, 11–20.
- Shoemaker, P. J., & Vos, T. (2009). *Gatekeeping Theory*. Taylor & Francis Group.
<http://ebookcentral.proquest.com/lib/ugalib/detail.action?docID=446721>
- Sidoti, O., & Dawson, W. (2024). *Social Media Fact Sheet*.
<https://www.pewresearch.org/internet/fact-sheet/social-media/>
- Small, M. L., & Adler, L. (2019). *The Role of Space in the Formation of Social Ties*.
- Smith, M. A., & Kollock, P. (1999). *Communities in Cyberspace*.
- Soliman, A., Hafer, J., & Lemmerich, F. (2019). A Characterization of Political Communities on Reddit. *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, 259–263. <https://doi.org/10.1145/3342220.3343662>
- Song, F. W. (2009). *Virtual Communities: Bowling Alone, Online Together*. Peter Lang.
- Squirrell, T. (2019). *Platform dialectics: The relationships between volunteer moderators and end users on reddit*. <https://doi.org/10.1177/1461444819834317>

- Stano, S. (2023). Linguistic guerrilla warfare 2.0: On the “forms” of online resistance. *Rivista Italiana Di Filosofia Del Linguaggio*, 2022: *SFL-Language, powers, rights* (eds. A. Bertollini & S. Garelli). <https://doi.org/10.4396/2022SFL13>
- Stanoevska-Slabeva, K., Sacco, V., & Giardina, M. (2012). Content Curation: A new form of gatewatching for social media. *Documento Electrónico. Recuperado El*, 16. <https://www.academia.edu/download/34329865/Katarina.pdf>
- Steen, E., Yurechko, K., & Klug, D. (2023). You Can (Not) Say What You Want: Using Algospeak to Contest and Evade Algorithmic Content Moderation on TikTok. *Social Media + Society*, 9(3), 20563051231194586. <https://doi.org/10.1177/20563051231194586>
- Stratton Oakmont, Inc. v. Prodigy Servs (N.Y. Supreme Court May 24, 1995).
- Straub-Cook, P. (2018). Source, Please?: A content analysis of links posted in discussions of public affairs on Reddit. *Digital Journalism*, 6(10), 1314–1332. <https://doi.org/10.1080/21670811.2017.1412801>
- Stryker, S., Owens, T. J., & White, R. W. (2000). *Self, Identity, and Social Movements*. U of Minnesota Press.
- Suzor, N. P. (2019). *Lawless: The Secret Rules That Govern Our Digital Lives*. Cambridge University Press.
- Tamí-Maury, I., Brown, L., Lapham, H., & Chang, S. (2017). Community-based participatory research through virtual communities. *Journal of Communication in Healthcare*, 10(3), 188–194. <https://doi.org/10.1080/17538068.2017.1337604>
- Tannen, D., & Trester, A. M. (Eds.). (2013). Discourse in Web 2.0: Familiar. In *Discourse 2.0: Language and New Media*. Georgetown University Press.

- Team EMB. (2023). *The Vital Role of Content Moderation: A Deep Dive into Online Safety*. EMB Global. <https://blog.emb.global/vital-role-of-content-moderation/>
- Tech Against Terrorism. (2023). *The Online Regulation Series 3.0: The Handbook*. Tech Against Terrorism. <https://techagainstterrorism.org/hubfs/Online%20Regulation%20Series%203-0-%20Tech%20Against%20Terrorism.pdf>
- The most important city in America: R/philadelphia*. (2024, November 4). [Online post]. https://www.reddit.com/r/philadelphia/comments/1gje9r6/the_most_important_city_in_america/
- Tobi, A. (2024). Towards an Epistemic Compass for Online Content Moderation. *Philosophy & Technology*, 37(3), 109. <https://doi.org/10.1007/s13347-024-00791-3>
- Tönnies, F. (1887). *Gemeinschaft und Gesellschaft: Abhandlung des Communismus und des Socialismus als empirischer Culturformen*. Fues. <https://archive.org/details/gemeinschaftund00tngoog>
- Tufekci, Z. (2017). *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press.
- Üblacker, J., Liebig, S., & Hamad, H. (2024). Catalysts of connection. The role of digital information and communication technology in fostering neighbourhood social cohesion: A systematic review of empirical findings. *Urban Studies*, 61(16), 3167–3186. <https://doi.org/10.1177/00420980241281502>
- Udupa, S., Maronikolakis, A., & Wisiolek, A. (2023). Ethical scaling for content moderation: Extreme speech and the (in)significance of artificial intelligence. *Big Data & Society*, 10(1), 20539517231172424. <https://doi.org/10.1177/20539517231172424>

- Usher, N. (2021). *News for the Rich, White, and Blue: How Place and Power Distort American Journalism*. Columbia University Press.
- Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, 15(3), 398–405. <https://doi.org/10.1111/nhs.12048>
- Van Dijck, J. (2009). Users like you? Theorizing agency in user-generated content. *Media, Culture & Society*, 31(1), 41–58. <https://doi.org/10.1177/0163443708098245>
- Veglis, A. (2014). Moderation Techniques for Social Media Content. In G. Meiselwitz (Ed.), *Social Computing and Social Media* (Vol. 8531, pp. 137–150). Springer International Publishing. <https://doi.org/10.1007/978-3-319-07632-4>
- Vos, K. (2005). Community development policy and legislation: The Dutch case. *Community Development Journal*, 40(4), 405–418. <https://doi.org/10.1093/cdj/bsi085>
- Vu, H. T. (2014). The online audience as gatekeeper: The influence of reader metrics on news editorial selection. *Journalism*, 15(8), 1094–1110. <https://doi.org/10.1177/1464884913504259>
- Wang, H. M., Bulat, B., Fujimoto, S., & Frey, S. (2022). Governing for Free: Rule Process Effects on Reddit Moderator Motivations. In C. Stephanidis, M. Antona, S. Ntoa, & G. Salvendy (Eds.), *HCI International 2022 – Late Breaking Posters* (pp. 97–105). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-19682-9_14
- Wang, W. (2019). *A Study of Digitally Enhanced People–Space Interaction: A Place-Centric Perspective*. <https://doi.org/10.1177/1206331219881352>
- Wasike, B. S. (2011). *Framing Social News Sites: An Analysis of the Top Ranked Stories on Reddit and Digg*.

- Webber, D. J., Hughes, E., Pacheco, G., & Parry, G. (2022). Investment in digital infrastructure: Why and for whom? *REGION*, 9(1), 147–163. <https://doi.org/10.18335/region.v9i1.415>
- Welbers, K., & Opgenhaffen, M. (2018). Social media gatekeeping: An analysis of the gatekeeping influence of newspapers' public Facebook pages. *New Media & Society*, 20(12), 4728–4747. <https://doi.org/10.1177/1461444818784302>
- Weld, G., Zhang, A. X., & Althoff, T. (2024). Making Online Communities 'Better': A Taxonomy of Community Values on Reddit. *Proceedings of the International AAAI Conference on Web and Social Media*, 18, 1611–1633. <https://doi.org/10.1609/icwsm.v18i1.31413>
- Wellman, B., Boase, J., & Chen, W. (2002). The networked nature of community: Online and offline. *It & Society*, 1(1), 151–165.
- West, S. R., Peters, J., & Anastasopoulos, L. J. (2023). All the News That's Fit to Be Identified: Facilitating Access to High Quality News through Internet Platforms. *Colorado Technology Law Journal*, 21(1), 1–20.
- What to do about the automoderator?* (2024, December 12). [Reddit Post]. R/Sanfrancisco. www.reddit.com/r/sanfrancisco/comments/1hcalpw/what_to_do_about_the_automoderator/
- Why are Scott Wiener's posts being banned here?* (2024, November 19). [Reddit Post]. R/Sanfrancisco. www.reddit.com/r/sanfrancisco/comments/1gv1awg/why_are_scott_wieners_posts_being_banned_here/

- Wiggins, B. E. (2019). *The Discursive Power of Memes in Digital Culture | Ideology, Semiotics*.
<https://www.taylorfrancis.com/books/mono/10.4324/9780429492303/discursive-power-memes-digital-culture-bradley-wiggins>
- Willson, M. A. (2006). *Technically together: Re-thinking community within techno-society*. Peter Lang.
- Wilms, L. K., Gerl, K., Stoll, A., & Ziegele, M. (2024). Technology acceptance and transparency demands for toxic language classification – interviews with moderators of public online discussion fora. *Human–Computer Interaction*, 0(0), 1–26.
<https://doi.org/10.1080/07370024.2024.2307610>
- X Corp. vs. Robert A. Bonta: Hearing on 2:23-CV-01939-WBS-AC*, United States District Court for the Eastern District of California Sacramento Division (2025).
- Zahnow, R. (2024). Place type or place function: What matters for place attachment? *American Journal of Community Psychology*, 73(3–4), 446–460. <https://doi.org/10.1002/ajcp.12722>
- Zannettou, S., Bradlyn, B., De Cristofaro, E., Kwak, H., Sirivianos, M., Stringini, G., & Blackburn, J. (2018). What is Gab: A Bastion of Free Speech or an Alt-Right Echo Chamber. *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, 1007–1014. <https://doi.org/10.1145/3184558.3191531>
- Zimmer, M. (2017). “But the data is already public”: On the ethics of research in Facebook. In K. Miller & M. Taddeo (Eds.), *The ethics of information technologies*. Routledge.